SAD PHASING OF PROTEINS USING XENON GAS

A Thesis Submitted to the College of
Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in the Department of Pharmacy & Nutrition
University of Saskatchewan
Saskatoon

By

Brennan D. Bonnet

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professors who supervised my thesis work, or in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any materials in my thesis. Requests for permission to copy or make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Pharmacy & Nutrition

University of Saskatchewan

Saskatoon, Saskatchewan, S7N 5E5

Abstract

Structural biology is a branch of science related to biochemistry, biophysics, and molecular biology that deals with the molecular structures of biological macromolecules, in particular nucleic acids and proteins. Structure-guided drug design uses three-dimensional knowledge of protein structures to design small molecules which block the action of specific proteins. When crystals of theses macromolecules and their complexes can be obtained, their crystal structures can be determined by using isomorphous differences between a native structure and a derivative structure. This allows crystallographers to determine the coordinates of a small number of heavy atoms which provide initial phases for macromolecules. The advent of synchrotron radiation allowed determination of a heavy atom substructure by use of anomalous differences using either multiple wavelengths (MAD) or a single wavelength (SAD); the latter has become the most common phasing method in crystallography and is the method used in this study. The use of SeMet has been by far the most successful method employed in SAD. However, in some cases production of SeMet proteins is not possible thus necessitating additional options, for example, xenon.

Noble gases such as xenon may be used in SAD experiments by binding to various, non-specific sites. Advances in noble gas pressurization systems like the Hampton Research Xenon Chamber have greatly eased the production of noble gas derivatives, xenon itself being a prime candidate with a very strong anomalous signal when compared to lighter noble gases like krypton and argon. Investigation of the phasing properties of xenon was carried out on test proteins hen egg white lysozyme (HEWL), thermolysin, glucose isomerase, and thaumatin II. Phases were successfully determined for all four proteins including thaumatin II which did not bind xenon but was successful due to the anomalous signal from 17 native sulfurs. The three remaining proteins showed varying occupancies and numbers of sites including xenon sites in thermolysin and glucose isomerase which have not been observed previously. This document will serve as a guide for the preparation of xenon derivative crystals and provides a strategy for the collection and processing of data from xenon derivatives.

Acknowledgements

I would like to take a moment to acknowledge all those in my life without whom I would not have had this great opportunity. I would like to thank my supervisor Dr. Pawel Grochulski for offering me this wonderful opportunity to work and gain experience at the Canadian Light Source. His expertise and character were essential to the success of my project. The countless hours spent on my behalf reflect my own and it has been a privilege to study protein crystallography under his guidance. Thanks to my committee members Dr. Michel Fodje and Dr. David Sanders for taking on the responsibility of preparing and directing me. Special thanks to Dr. Fodje for his repeated help as a "Linux Ninja".

I would like to thank all of the institutions that made this opportunity possible. For financial support I would like to thank the Molecular Design Research Group (MDRG) and the Canadian Light Source (CLS). Thanks to the College of Pharmacy & Nutrition for my education and thanks to the University of Saskatchewan, in particular, for the pursuit of innovation in science and technology that led to the building of Canada's synchrotron facility in my home city of Saskatoon.

Thanks once again to the Canadian Light Source (CLS) and the Canadian Macromolecular Crystallography Facility (CMCF) for the warm welcome. Thanks to all of the members of CMCF Dr. Pawel Grochulski, Dr. Michel Fodje, Dr. Shaun Labiuk, Dr. Julien Cotelesage, James Gorin, Kathryn Janzen, and Osama Alaidi for their help with anything and everything. These are the people from whom I acquired all of the knowledge and skills necessary to successfully complete my project; their help has been invaluable.

And last but not least, I would like to thank my family. A huge thanks to my parents Val and Maurice for their sacrifices and for reasons beyond number and scope; I would not be where I am if not for them. Thanks to my sisters Danielle and Celeste, to my niece Taylor, to my nephew Jordan, and to the rest of my family and friends for understanding that this project has really cut into our time together and for continuing to have me around even though sometimes I talk nonsense about crystals and synchrotrons.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

SAD	Single-Wavelength Anomalous Dispersion
S-SAD	Sulfur-SAD
MAD	Multiple-Wavelength Anomalous Dispersion
CMCF	Canadian Macromolecular Crystallography Facility
CLS	Canadian Light Source
MIR	Multiple Isomorphous Replacement
SeMet	Selenomethionine
PDB	Protein Data Bank
HEWL	Hen Egg White Lysozyme
BM	Bending Magnet
MxDC	Macromolecular Crystallography Data Collection
MX	Macromolecular Crystallography
LSQ	Least Squares Fit
ASU	Asymmetric Unit
DCM	Double Crystal Monochromator
SGU	Small Gap Undulator
PEG	Polyethylene Glycol
CSD	Cambridge Structural Database
SGDD	Structure-Guided Drug Design

NMR	Nuclear Magnetic Resonance

1.0 Hypothesis & Objectives

1.1 Hypothesis

Xenon derivatives produced by xenon gas pressurization can be used in SAD experiments for phasing of macromolecules by as few as one partially occupied xenon site.

1.2 Objectives

1. Develop a reliable, step-by-step procedure which will allow phasing of *de novo* proteins using the anomalous signal from bound xenon atoms. The procedure will include:
    - Gas pressure and pressurization time
    - Data collection strategy
    - Data processing strategy
2. Test the proposed procedure on a protein which is known to bind xenon.
    - Grow HEWL crystals
    - Derivatize HEWL crystals using Hampton Research Xenon Chamber
    - Collect Data
    - Identify anomalous signal
    - Locate anomalous substructure using anomalous differences and build model
    - Refinement of f'' values
    - Visual inspection of xenon sites
    - Compare model with Protein Data Bank (PDB)
3. Use the procedure for efficient phasing of several other test proteins.
    - Thaumatin II, thermolysin, and glucose isomerase

2.0 Introduction

Macromolecular crystallography is a main method for obtaining high resolution structures of macromolecules such as proteins, viruses, nucleic acids (DNA and RNA), and carbohydrates. This high resolution helps in explaining the detailed mechanisms by which these macromolecules carry out their functions in living cells and organisms. This year the Protein Data Bank (PDB) has surpassed 100,000 entries with about 80% of these structures determined at synchrotrons. The dominating method for phasing has made use of selenomethionine-SAD[1]. However, selenomethionine can be toxic and incorporation may result in very low expression yields.

In such cases, or when it is not possible to produce selenomethionine-containing proteins or to grow their crystals, an alternative method can be used: sulfur-SAD (S-SAD). However, this method normally requires at least one sulfur per approximately 25 residues in the protein sequence[2]. Xenon derivatization is a method that would be very beneficial for cases when there are not enough sulfurs, but could also be recommended as an addition to any S-SAD experiment in order to increase anomalous signal and improve the chances of solving the crystal structure.

2.1 Synchrotron Radiation

Synchrotron radiation is produced when charged particles such as electrons are forced to accelerate at relativistic speeds. Large machines are built to force electrons to travel in a circular path at near the speed of light. This produces a very narrow and intense beam of polychromatic radiation. The use of double crystal monochromators (DCM) allows researchers to tune synchrotron radiation to a desired monochromatic wavelength. Synchrotron radiation is essential to modern protein x-ray crystallography whereby very weak anomalous signals originating from anomalously scattering atoms are measured and exploited. Choosing the proper wavelength is important for maximizing anomalous signal. Furthermore, the high intensity of synchrotron radiation allows data collection in a fraction of the time required with conventional x-ray sources[3].

3rd generation synchrotron facilities such as the Canadian Light Source (CLS) provide further improvement in radiation intensity allowing for even faster data collection. This generation of synchrotrons is made possible with the advent of insertion devices called wigglers and undulators. An insertion device is a periodic magnetic structure which causes electrons to travel in a sinusoidal path as they pass through the device gap; the result of this motion is an increase in brilliance. Undulators are specially tuned to produce constructive interference between electrons resulting in a narrow radiation bandwidth with an increase in brilliance superior to wigglers[3].

2.1.1 CLS

The Canadian Light Source located in Saskatoon, Saskatchewan at the University of Saskatchewan is Canada's national synchrotron research facility. This state of the art 3rd generation synchrotron is dedicated to advancing Canadian science and providing Canadian industry with a world class facility for conducting academic and industrial research. The staff of the Canadian Macromolecular Crystallography Facility operates two crystallography beamlines CMCF-ID (08ID-1) and CMCF-BM (08B1-1) providing on site, mail-in, and remote data collection services[4].

2.2 Methods of Protein Crystallography

Protein crystallography takes advantage of the fact that proteins grow into crystals under certain conditions. These protein crystals are made up of regular, repeating copies of the simplest unit called the asymmetric unit (ASU); several copies of the ASU may be present in the unit cell. Unit cells may be arranged into fourteen Bravais lattices and when combined with all 32 point groups, there are 230 possible space groups when translational symmetry is considered. Of the 230 possible space groups, only those which preserve chirality are valid which reduces the number of possibilities to 65[5].

The ordered, repeating nature of crystals results in both diffraction and amplification of scattered x-rays. The tiny fraction of x-ray diffraction coming from each copy of the ASU is amplified by the number of copies in the crystal. The resulting signal is sampled onto

discrete points and is large enough to measure. Synchrotron radiation is almost always preferable due to high intensity and high collimation both of which allow shorter exposure times and quicker experiments with increased accuracy. Synchrotron radiation also has the advantage of being tuneable. This allows the energy at which experiments are carried out to be changed as in the case of SAD/MAD experiments.

2.2.1 X-ray Diffraction

X-rays in the energy range appropriate for diffraction experiments interact with matter through their electrons. The atom's electrons first absorb the x-ray energy and then act as an elastic point source by re-emitting x-rays without changing their energy. This happens with each and every atom in the crystal producing wave interference; essentially, protein crystals are three-dimensional diffraction gratings for electromagnetic radiation in the x-ray range[5]. Wave interference manifests as a diffraction pattern of constructive x-ray reflections and represents a snapshot of the protein in reciprocal space. If enough snapshots are taken and reflections measured, the electron density of the atoms which produced the diffraction patterns may be reconstructed[5].

A crystal can be modeled as a set of parallel planes with constant interplanar distance S(hkl). The indices h, k, and l describe the family of planes which cut the unit cell into h, k, and l equal parts e.g. hkl=201 means that this family of planes cuts the unit cell on the x-axis into 2 pieces each a/2 in length, does not cut the unit cell on the y-axis and so runs parallel to the y-axis, and cuts the unit cell along the z-axis into 1 piece of length c[5]. For a given wavelength, λ, angles of incidence and reflection θ with respect to the planes (Figure 2.1), and some integer, n=1,2,3,…, it is possible to calculate where constructive interference will occur:

$$2S(hkl)\sin\theta = n\lambda \qquad (2.1)$$

Bragg's law (2.1) describes the conditions for constructive interference[5].

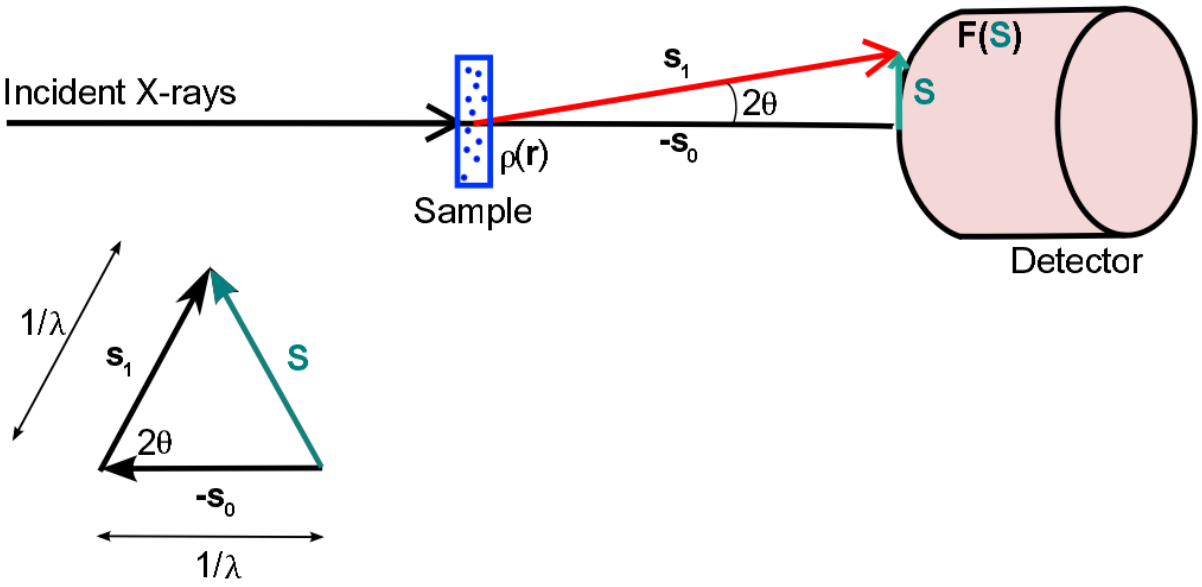


Figure 2.1 Definition of the scattering vector S

The scattering vector **S** is the difference between the incident beam $\mathbf{s_0}$ and the scattered beam $\mathbf{s_1}$.

Every point in the protein electron density, ρ(r), produces a partial wave which is described by a wave of the form exp(iφ) with a relative phase of $\varphi=2\pi\mathbf{S}\cdot\mathbf{r}$ and a magnitude based on the electron density within the volume element. After integration over the atomic volume, the total resulting wave from an atom is:

$$f = \int_{r}^{V\,(atom)} \rho(\mathbf{r}) \exp(2\pi i \mathbf{S}\cdot\mathbf{r})\, d\mathbf{r} \qquad (2.2)$$

From equation 2.2, an individual atom's ability to scatter x-rays is dependent on the electron density ρ at **r**, and on the scattered wave vector, **S**, which is the difference between the incoming and scattered wave vectors $\mathbf{s_0}$ and $\mathbf{s_1}$, respectively (Figure 2.1). **S** is dependent on the type of atom doing the scattering and falls off with angle θ as $|\mathbf{S}|=(2\sin\theta)/\lambda$. The scattering factor for an entire molecule then is simply the summation of the individual scattering factors from each point, **r**:

$$\mathbf{F}(\mathbf{S}) = \sum_{j=1}^{n\ atoms} f_j \exp(2\pi i \mathbf{r}_j \cdot \mathbf{S}) \qquad (2.3)$$

where $\mathbf{F}(\mathbf{S})$ is a vector in reciprocal space and $\mathbf{F}(\mathbf{S}) \rightarrow \mathbf{F}(hkl)$ with the scalar product $\mathbf{r}_j \cdot \mathbf{S} = hx + ky + lz$.

As mentioned earlier, a diffraction image is an image of the protein in reciprocal space. Each measured reflection is called a structure factor, $\mathbf{F}(hkl)$, and each contains within it all of the structural information for a whole family of parallel planes hkl. As the crystal is rotated and images are taken at different orientations, equation 2.3 is satisfied for some reflections which begin to interfere constructively while other reflections disappear. It is possible to continue in this manner until each reflection has been recorded several times. The electron density in direct space is related to the structure factors in reciprocal space by a Fourier transform.

$$\rho(xyz) = FT[\mathbf{F}(hkl)] = \frac{1}{V} \iiint_{-h-k-l}^{hkl} \mathbf{F}(hkl) \exp[-2\pi i(hx + ky + lz)] dh dk dl \quad (2.4)$$

In practice, data are not continuous but are instead produced at discrete points and equation 2.4 reduces to a triple Fourier summation using the relationship $\mathbf{F}=|\mathbf{F}|e^{-i\alpha}$:

$$\rho(xyz) = \frac{1}{V} \sum_{-h-k-l}^{hkl} |\mathbf{F}(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)] \qquad (2.5)$$

where V is the volume of the unit cell, $|\mathbf{F}(hkl)|$ is the structure factor amplitude arising from planes having indices hkl, $\alpha(hkl)$ is the phase, and xyz are the coordinates of the electron density, $\rho(xyz)$, in direct space. In order to calculate the electron density in direct space, $\rho(xyz)$, all structure factors must be measured and then applied to all points in space. Structure factors with higher indices are more difficult to measure and correspond to more closely spaced planes. The closer planes are to one another, the more finely sampled is the crystal lattice leading to higher resolution $\rho(xyz)$[5].

Structure factors can be represented by complex numbers on the complex (Argand) plane meaning that they have both amplitude and a phase which must be determined. Structure factor amplitudes are easily measured from diffraction images because diffraction intensity is proportional to the square of the structure factor amplitude:

$$I(hkl) \propto |\mathbf{F}(hkl)|^2 \qquad (2.6)$$

Phases, α(hkl), cannot be measured directly. Instead, other methods are used to obtain phases; this is called the phase problem in crystallography[5].

2.2.2 Derivatization

Derivatization is the process by which phase information is obtained by incorporating one or more heavy atoms into a native protein crystal. The phase problem has long been solved using isomorphous differences between a native and one or more derivative data sets. Using this method, the positions of the heavy atoms can be approximated using the difference between the derivative and native datasets (Figure 2.2). An extension of the isomorphous difference method, and the one used in this study, is to measure the very small anomalous differences produced when atoms reveal anomalous x-ray scattering. This method relies on the fact that chosen derivative atoms exhibit significant anomalous scattering in diffraction experiments while native protein atoms such as carbon and nitrogen do not. More recently, the anomalous signal coming from native sulfur atoms in methionine and cysteine residues have provided sufficient anomalous signal to solve structures[6–8]. Making use of anomalous differences is a much more sensitive experiment and is made possible due to advances in protein crystallography, software algorithms, and synchrotron techniques. The current standard for *de novo* phasing done via SeMet labeled protein requires labor-intense additional expression, purification, crystallization steps, and data collection whereas derivatization with xenon uses existing native crystals.

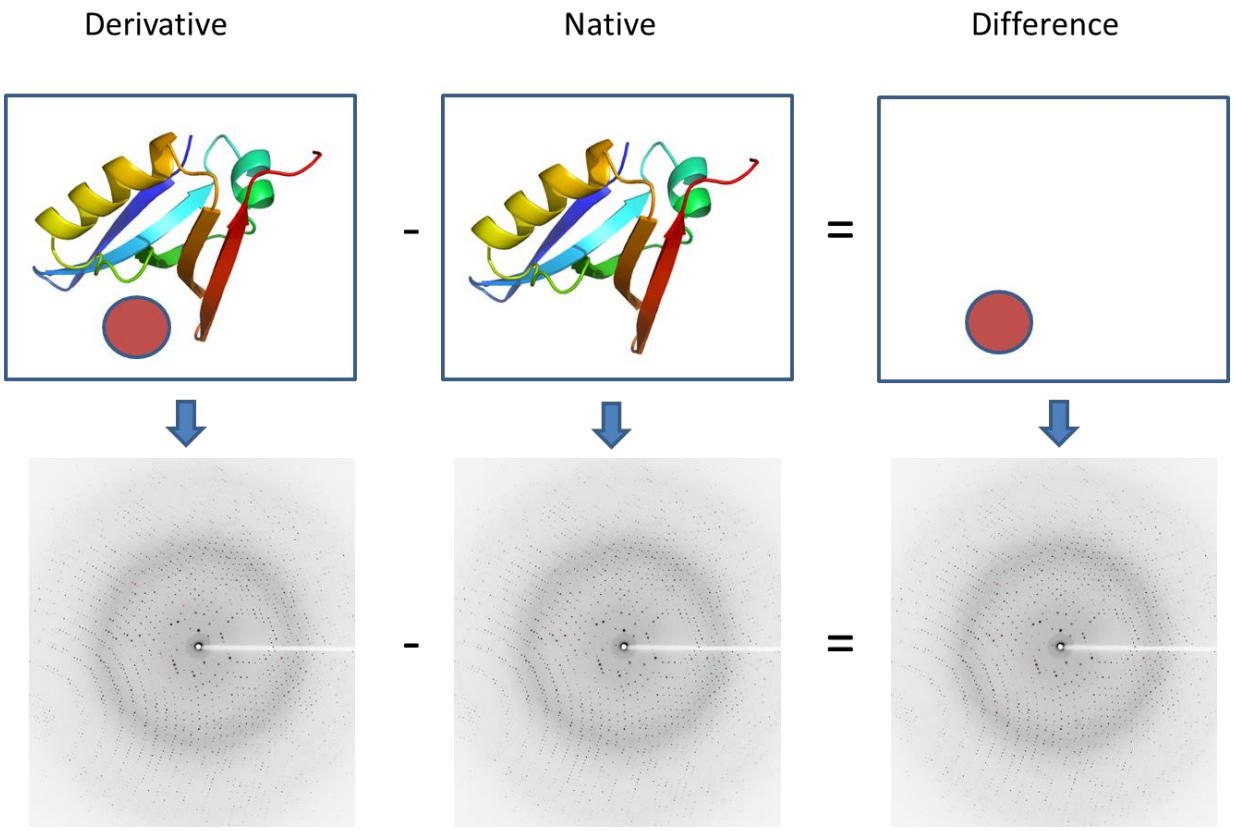


Figure 2.2 Protein derivatization

Principle of difference method used in protein crystallography to solve the phase problem. The experimental phasing relies on isomorphous difference data. It can originate from heavy atoms (SIR, MIR), anomalous differences between Bijovet pairs of the same data set (SAD), dispersive and anomalous differences between different wavelengths (MAD), or any combination thereof.

2.2.2.1 Direct Methods

Direct methods were initially developed to solve small molecule crystal structures and are good for solving structures up to 100 non-hydrogen atoms[9]. These methods are impractical for determining protein structures themselves but are perfect for determining the coordinates of a small heavy atom substructure and consequently the initial phases for a protein. These methods make use of mathematical relationships which exist between combinations of phases. Once a few phases can be given approximate starting values or are

known, these relationships can be used to "bootstrap" additional phases to build a complete set[10,11].

Direct methods require comparison between reflections. Since structure factor amplitudes fall off with the scattering angle, θ, they must be normalized using the expectation value, ‹F^2›, for that scattering angle:

$$E_0^2 = \frac{F_0^2}{\langle F^2 \rangle} \qquad (2.7)$$

$$\langle F^2 \rangle = k\varepsilon \sum F_i^2 \exp\left(-\frac{2B \sin^2 \theta}{\lambda^2}\right) \qquad (2.8)$$

where λ is the x-ray wavelength, B is the atomic displacement factor, k is the scale factor, and ε is a symmetry enhancement factor. In the calculation of ‹F^2› it is important that all reflections are measured in the relevant θ range including the very weak reflections[11].

An important relationship which exists between structure factors was formulated by Sayre[12] and is based on two assumptions. The first criterion is that the electron density in a structure can never be negative, called positivity, and the second is that atoms are well-defined points of electron density, called atomicity. Based on these assumptions, we can calculate the structure factor of some hkl reflection as the sum of the products of all pairs of reflections whose indices add up to hkl:

$$F_{hkl} = k \sum_{h'k'l'} F(h'k'l') \cdot F(h - h', k - k', l - l') \qquad (2.9)$$

For example, reflection (321):

$$E(321) = E(100) \cdot (E221) + E(110) \cdot E(211) + (E111) \cdot E(210) + \cdots \qquad (2.10)$$

This relationship may not seem particularly useful but it gives very good approximations because any pairs of reflections in 2.10 where one or both reflections are weak can initially be ignored. However, in the case that both structure factors are strong and the structure factor being calculated is also strong, then there is a good chance that the structure factor being calculated from 2.10 is accurate[11].

Another relationship called the triplet relationship (2.11) relates the phases and indices of 3 reflections:

$$\alpha(hkl) + \alpha(h'k'l') \cong \alpha(h + h', k + k', l + l') \qquad (2.11)$$

For example, reflection (022):

$$\alpha(213) + \alpha(-21-1) \cong \alpha(022) \qquad (2.12)$$

For a system that is small enough to allow a large percentage of these relationships to be examined at once, it is possible to build enough constraints on phases in order to obtain fairly good initial estimates. One more important relationship for relating phases is the tangent formula derived by Karle & Hauptman[10]:

$$\tan \alpha_h = \frac{\sum_k |E(k)\cdot E(h-k)| \sin(\alpha(h)+\alpha(h-k))}{\sum_k |E(k)\cdot E(h-k)| \cos(\alpha(h)+\alpha(h-k))} \qquad (2.13)$$

2.2.2.2 Patterson Method

The other method for locating heavy atoms and obtaining initial phases is by use of the Patterson function

$$P(uvw) = \frac{1}{V}\sum_{hkl} |\mathbf{F}(hkl)|^2 \cos[2\pi(hu + kv + lw)] \qquad (2.14)$$

where uvw are coordinates in Patterson space which are analogous to atomic coordinates xyz except that they correspond to interatomic vector peaks[5,11]. This means that some peaks are the result of self-vectors and will lie on the origin of the Patterson map. Patterson peaks are also centrosymmetric meaning that, given two atoms, the peak arising from 2-1 is centrosymmetrically related to the peak arising from 1-2. For N atoms, the Patterson map will contain N origin peaks with N^2-N off-origin peaks (Figure 2.3).

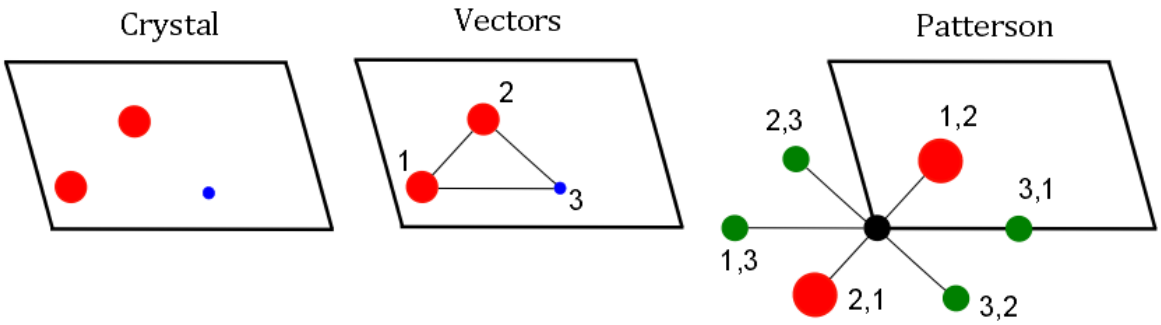


Figure 2.3 Patterson maps

Each peak in the Patterson map corresponds to an interatomic vector between atoms in a crystal. An increase in the number of atoms quickly yields a Patterson map which has too many peaks to interpret.

Due to the large number of peaks, it is not possible to determine the locations of atoms directly and this problem becomes much worse as the number of atoms increases. However, the Patterson function can still be of use as it is useful for determining the locations of a small heavy atoms substructure. The advantage of using the Patterson function is that $|\mathbf{F}(hkl)|$ values can be measured directly from diffraction images by relation 2.6 without any need for phase information[5,11,13].

The intensity of a Patterson peak is estimated as the product of the number of electrons that the two atoms have. In this manner, peaks corresponding to pairs of "light" atoms will be less intense than peaks corresponding to heavy atom light atom pairs which will be less intense than peaks arising from pairs of heavy atoms. Knowing this about Patterson maps, it is easy to see that the most intense peaks will come as a result of heavy atom relationships. This allows determination of the locations of a few heavy atoms by looking at the most intense peaks in the Patterson map. Harker sections (peaks, lines, and planes) are special locations on the Patterson map where peaks arising from symmetry-related atoms occur (Figure 2.4). The difference between an atomic position x,y,z and a symmetry related position such as -x,½+y,-z ($P2_1$) provides the locations of these Harker sections.

Once these sections are found, then the most intense peaks in these sections correspond to symmetry-related heavy atoms, the locations of which can be easily determined[5,11].

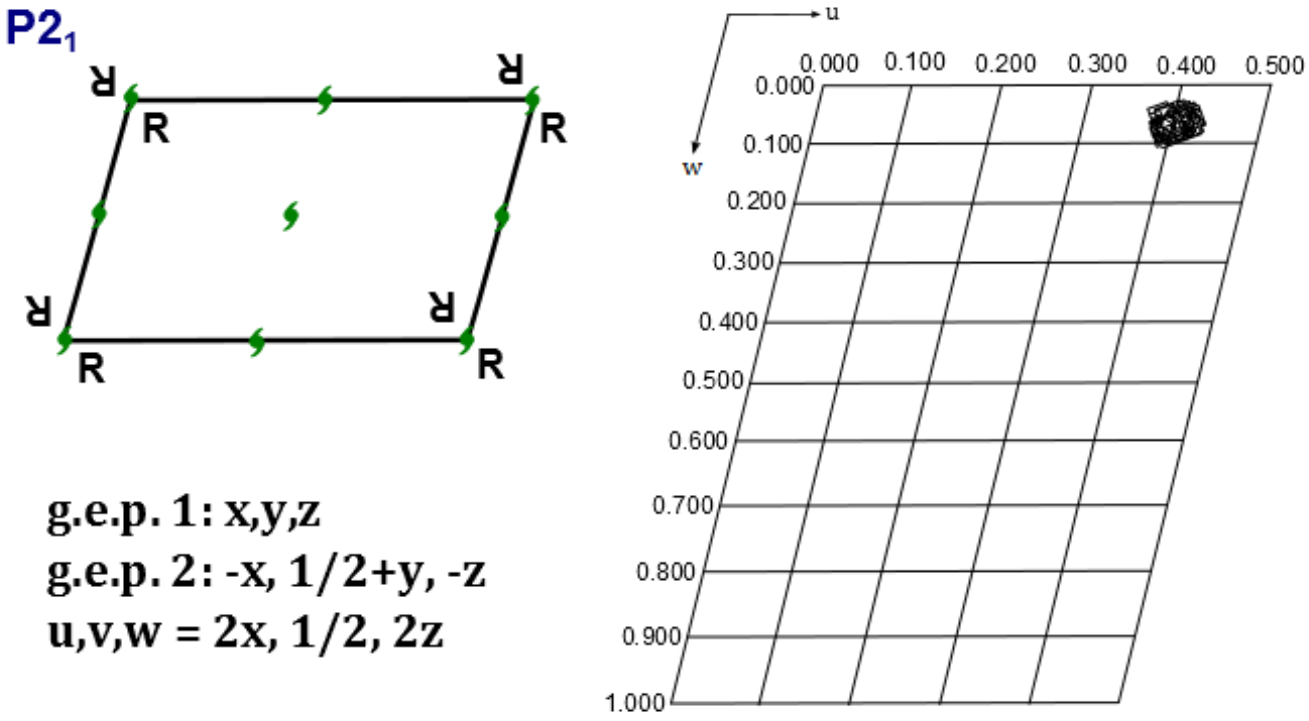


Figure 2.4 Harker vectors, sections

Patterson peaks occur at the endpoints of vectors between all pairs of atoms in a unit cell of the crystal. Consider one site and its general equivalent positions (g.e.p.'s) related by space group symmetry; vectors between these symmetry-related atoms are known as *Harker vectors*. For space group $P2_1$ (2_1 axis along b), xyz are the site coordinates, and uvw are the Patterson coordinates. In $P2_1$ the *Harker vector* always occurs in the section defined by v=1/2 regardless of the site coordinates. This section v=1/2 in the Patterson for space group $P2_1$ is a *Harker section*.

2.2.1 SAD (Single-Wavelength Anomalous Dispersion)

For years researchers have solved the phase problem by soaking native crystals in salt solution of heavy atoms or by co-crystallization in order to produce heavy atom derivatives. Historically, the seven most successful compounds have been K_2PtCl_4,

$KAu(CN)_2$, K_2HgI_4, $UO_2(C_2H_3O_2)$, $HgCl_2$, para-chloro mercury benzoic acid sulfonate (PCMBS) and $K_3UO_2F_5$ [14] and by producing several different derivatives, phase information may be obtained by a process known as Multiple Isomorphous Replacement (MIR). This condition of isomorphism is, in fact, not only a condition for MIR but is required for any phasing technique in which some heavy atom is introduced and multiple derivatives are required.

More recently, anomalous difference phasing methods take advantage of the atom's ability to absorb x-rays of certain wavelengths so called absorption edges. Usually when x-rays are diffracted by protein crystals, structure factors amplitudes measured from reflections on opposite sides of the same plane are equal:

$$|\mathbf{F}(hkl)| = |\mathbf{F}(-h-k-l)| \qquad (2.15)$$

this is known as Friedel's Law. When the incident x-ray wavelength approaches an atom's absorption edge, a fraction of the x-rays are absorbed by the atom and then re-emitted with altered phase meaning that Friedel's Law no longer applies[5,13]. The total structure factor, F, for an anomalously scattering atom is the sum of a wavelength-independent part, f_0, and the wavelength-dependent real and imaginary parts of the anomalous scatterer:

$$F(\lambda) = f_0 + f'(\lambda) + i\, f''(\lambda) \qquad (2.16)$$

where f' is the real scattering component which is in phase with f_0 and usually negative, and f'' is the imaginary scattering component which is always $\pi/2$ ahead of phase of the real parts f_0 and f' [5]. These small but measurable differences between Friedel pairs allow crystallographers to extract phase information from a single crystal by single-wavelength anomalous dispersion (SAD) or by using multiple wavelengths (MAD) if necessary.

Anomalous dispersion methods work because most of the atoms in proteins are hydrogen, oxygen, carbon, and nitrogen which are considered "light" and do not exhibit significant anomalous scattering at the wavelengths used in crystallography experiments. Some anomalous scattering atoms occur naturally such as sulfurs from cysteine and methionine amino acids or iron from oxygen-carrying proteins. Other times, proteins may require the introduction of anomalous scatterers by methods such as the heavy atom salt soak[14],

replacement of methionine amino acids by selenium-containing amino acids called selenomethionine[6,15], iodine soaking which has been found to bind to various, non-specific binding sites[16,17], or by pressurization of noble gases such as xenon or krypton[18].

None of these methods works 100% of the time, however. Aside from being somewhat harsh, heavy-atom salt soaks will not work if the protein has no binding sites for that particular heavy atom. Replacement of methionine by selenium-containing selenomethionine has made selenium by far the most used phasing element[15] but even this method will not work if the protein contains little or no methionine amino acids. Sometimes methionine amino acids or metal binding sites may be incorporated by site-directed mutagenesis[19] but this process involves changing the protein's native structure which may not be desirable. Having more options for obtaining phase information is always beneficial and will aid in avoiding research dead ends due to encountering proteins which resist established methods.

2.2.1.1 Optimizing data collection

The two main aspects required for optimizing SAD data collection are low resolution completeness and accuracy (Figure 2.5)[20]. Low resolution completeness of reflections from 20-3.5Å is important because these are the strong reflections which are used in Patterson and direct methods to determine initial phases as described earlier[21]. The most important factors when maximizing data completeness are total rotation range to ensure that each reflection is measured as many times as possible and rotation start position which depends on crystal orientation and symmetry. SAD phasing works because of small but measurable differences in *Friedel pairs*; this difference is called the *Bijovet ratio*, $\Delta F/F$, and is usually on the order of only a few percent. Therefore SAD phasing methods depend very much on high accuracy in order to measure these differences. Accuracy depends on several factors the most important of which are high multiplicity, fine slicing which gives better reflection profiles, and longer exposure times to minimize the effect of timing errors. One more key consideration is that when obtaining high multiplicity data, images near the end of data

collection should be inspected for signs of radiation damage which is detrimental to anomalous phasing techniques[20,22,23].

	Molecular Replacement	Anomalous Phasing	High-Resolution Refinement	Ligand Search
Accuracy	+	++++	++	++
Low-Resolution Completeness	+++	+++	++	++
Resolution	+	+	+++	++
Overall Completeness	++	++	++	++
Automation	++	+	++	+++

Figure 2.5 Carrying out an optimal experiment[20]

The relative importance of various aspects of data collection for different types of crystallography experiments. Accuracy and low-resolution completeness are very important for anomalous phasing experiments.

2.2.2 Cryoprotection

It is well known that protein crystals may be flash cooled by plunging them into liquid nitrogen. This technique has been used extensively in protein crystallography to transport and preserve crystals. More importantly, the practice of flash cooling protein crystals extends the lives of crystals during experiments by reducing the amount of radiation damage. This is especially beneficial when doses are expected to be high as in the case of synchrotron radiation or when a high accuracy data set is needed[24].

Solvent channels which exist in all protein crystals may account for more than half of the crystal volume. Upon flash cooling, water molecules in the solvent channels and in solvent surrounding the crystal form hexagonal ice which is less dense than water. This volume

expansion is undesirable as it can lead to crystal deformation, destruction of the crystal lattice, and even cracking crystals. Flash cooling crystals and dehydration by removal of surrounding water can aid in reducing these negative effects. The best solution is the use of cryoprotectants such as glycerol or PEGs either immediately before flash cooling or as an ingredient in the crystallization conditions[24].

2.3 Xenon

Xenon belongs to the category of "heavy atoms" used in protein crystallography. With 34 electrons, it is lighter than standard heavy atoms such as mercury, platinum, or uranium. Xenon has been known to bind to proteins since the 1950's[25–27] and since then has been used to solve structures first using isomorphous differences and more recently using anomalous differences[28–31]. Xenon derivatives have shown that xenon will bind to many different sites including intra- and inter-molecular sites, inaccessible cavities, and exposed pockets[18]. These sites may also be composed of several types of residues including aliphatic, aromatic, and polar residues[30]. Xenon has been found to displace water molecules that are well defined in native structures but generally prefers hydrophobic pockets many of which are buried within the protein and away from the solvent. These sites are distinct from metal-binding sites and are therefore complementary to other methods and derivatives[18,28]. Furthermore, the occupancy of these binding sites increases depending on xenon pressure and an increase in pressure may also increase the number of binding sites[32].

Binding happens slowly as gas pressure is applied and xenon diffuses into a crystal through solvent channels over a few minutes[18,33–35]. Xenon is able to reach buried pockets within proteins via solvent channels due to its high solubility in water[36–38]. However, xenon binding is only possible for as long as an external gas pressure is applied. This makes xenon binding non-covalent and reversible leaving the protein structure wholly unchanged. Diffusion of xenon leaving the crystal is also slow. This is convenient as it allows time to release the gas pressure and then "freeze" xenon into crystals using liquid nitrogen[33,39,40].

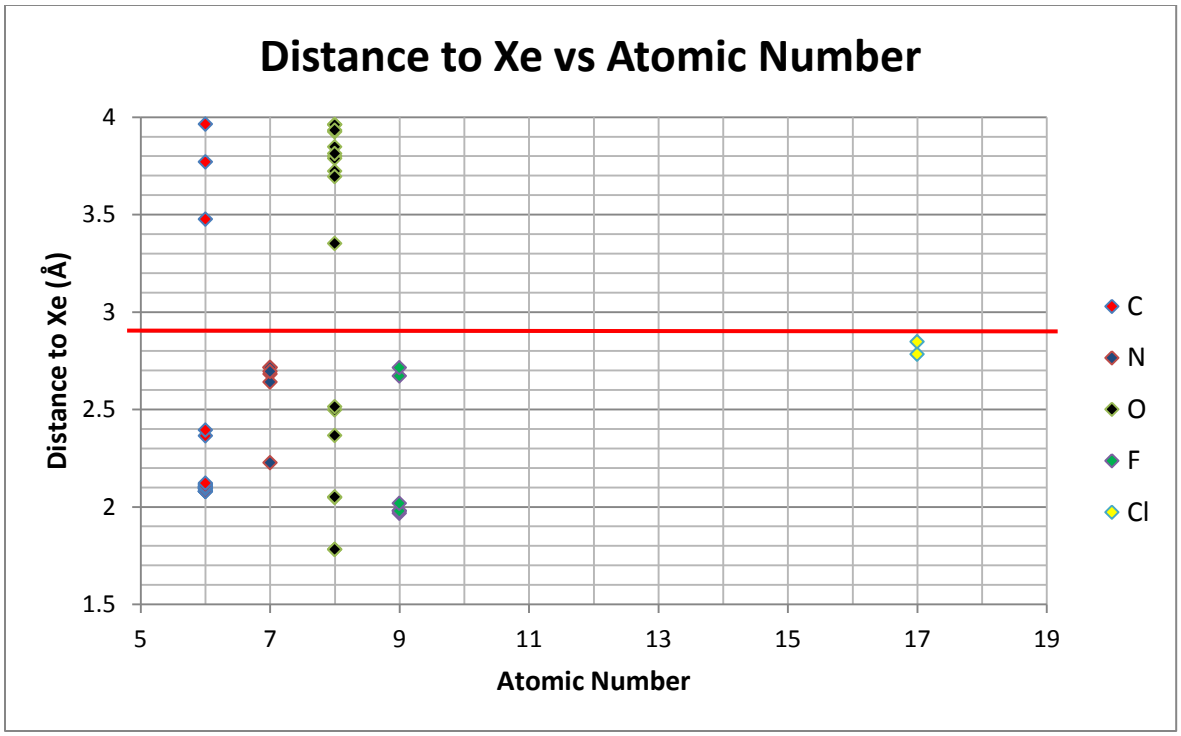


Figure 2.6 CSD statistics for xenon binding[41]

As a noble gas, xenon does not usually take part in chemical bonding. There are only 38 structures in the CSD which involve xenon. However, there are more than 80 compounds known[42]. Results of searching the CSD clearly demonstrate two groups of distances. Those less than 2.8Å represent covalent bonds between Xe and O, N, C, F and Cl atoms. Those larger than 3Å are described by 3 types of interactions: charged-induced dipole interactions, dipole-induced dipole interactions, and London interactions.

Xenon is a large, neutral atom which, for the most part, is unreactive and does not form bonds (Figure 2.6)[41,42]. Due to its large size, xenon has a polarizability α =4.00. This is relatively high for a single atom especially when compared to lighter noble gases such as krypton (α=2.47) and argon (α=1.63). High polarizability is the result of larger atoms and molecules having larger, more dispersed electron densities and is important for the types of interactions between noble gases and proteins. There are three types of interactions[43] which can occur between noble gas atoms and molecular groups in proteins as summarized by Schiltz *et al.*[18]:

1. *Charge-induced dipole interactions*: The induced dipole in the xenon or krypton atom is created by an external electric charge. These interactions could thus occur when a charged protein group is located in the vicinity of the noble gas atom.

$$U(r) \sim -\frac{q^2\alpha^2}{r^4} \qquad (2.17)$$

   The energy for this type of interaction is proportional to the square of the electronic polarizability of the noble gas, α, proportional to the square of the charge, q, and inversely proportional to the fourth power of the distance separating the center of the charge from the center of the noble gas atom, r. These are therefore potentially strong interactions (as compared with dipole-induced and London interactions) and are also active over larger distances.
2. *Dipole-induced dipole interactions*: The induced dipole in the xenon or krypton atom is created by an external electric dipole. These interactions occur when the noble gas atom is located in the vicinity of polar groups in the protein.

$$U(r) \sim -\frac{\mu^2\alpha}{r^6} \qquad (2.18)$$

   The binding energy involved in this type of interaction is proportional to the square of the magnitude of the dipole moment, μ, proportional to the polarizability of the noble gas atom, α, and inversely proportional to the sixth power of the distance separating the center of the dipole from the center of the gas atom, r. Hence, it is a truly short-range interaction.
3. *London interactions (also called dispersion interactions)*: These interactions exist between all molecules and atoms, even those that are uncharged and nonpolar. They are usually described as arising from the interaction of the instantaneous dipoles in both molecules (and atoms). Each molecule (and atom) possesses fluctuating dipoles according to the particular instantaneous distribution of the electrons. These instantaneous dipoles constantly change direction and

magnitude, each existing only for a minute fraction of time. However, the net overall interaction between these instantaneous dipoles is an attractive force:

$$U(r) \sim -\frac{I_1 I_2}{I_1 + I_2} \frac{\alpha_1 \alpha_2}{r^6} \qquad (2.19)$$

where I is the first ionization energy of each molecule. Once more, the interaction energy is proportional to the polarizability of the noble gas atom, α, (as well as being proportional to the polarizability of the other interacting atom, α) and inversely proportional to the sixth power of the distance separating the two atoms, r. These energies are again weak, usually only a few units of *k*T, but the xenon and krypton atoms in a binding site typically interact with several protein atoms, the interaction energies being added up. The binding of xenon and krypton to proteins is dominated by London interactions, mainly because the majority of binding sites are formed by nonpolar groups (aliphatic and/or aromatic side chains). But even if there are polar groups in a binding site, it can be shown that the London interactions generally contribute to a much larger extent to the binding than do the dipole-induced forces[18].

In SAD phasing, a strong anomalous signal, f'', is required. Xenon's L-I edge at 5.45keV has a very strong anomalous signal f''=13.4e. At energies this low, however, x-ray absorption due to interactions with air is significant and the anomalous signal is more difficult to measure accurately. That is, it may be lost in the measured data, but is still present in the crystal[44–47]. But xenon has a very strong anomalous signal even very far from any absorption edges e.g. f''=9.22e @ 7keV (Figure 2.7) which makes it a powerful phasing agent. A caveat is that the strong anomalous signal from a heavy atom such as xenon could appear weak due to there being too few binding sites or due to low xenon occupancies. The latter can cause difficulty when attempting to distinguish partially occupied xenon sites from other lighter anomalous atoms which also produce an anomalous signal at 7keV

(Figure2.8). Even so, SAD phasing using xenon has been successful using only a single half-occupied xenon site for proteins as large as 250kDa at 8keV and 650kDa at 5.4keV[34].

The anomalous signal from xenon is similar to iodine which has also been used in SAD studies to solve structures[48]. The lighter elements of the halogen and noble gas groups (Figure 2.9) have also been used to solve structures however their anomalous signal is relatively weak when compared to xenon and iodine. Obtaining sufficient anomalous signal from these lighter elements necessitates a large number of binding sites, the use of multiple wavelengths near an absorption edge (MAD), or both.

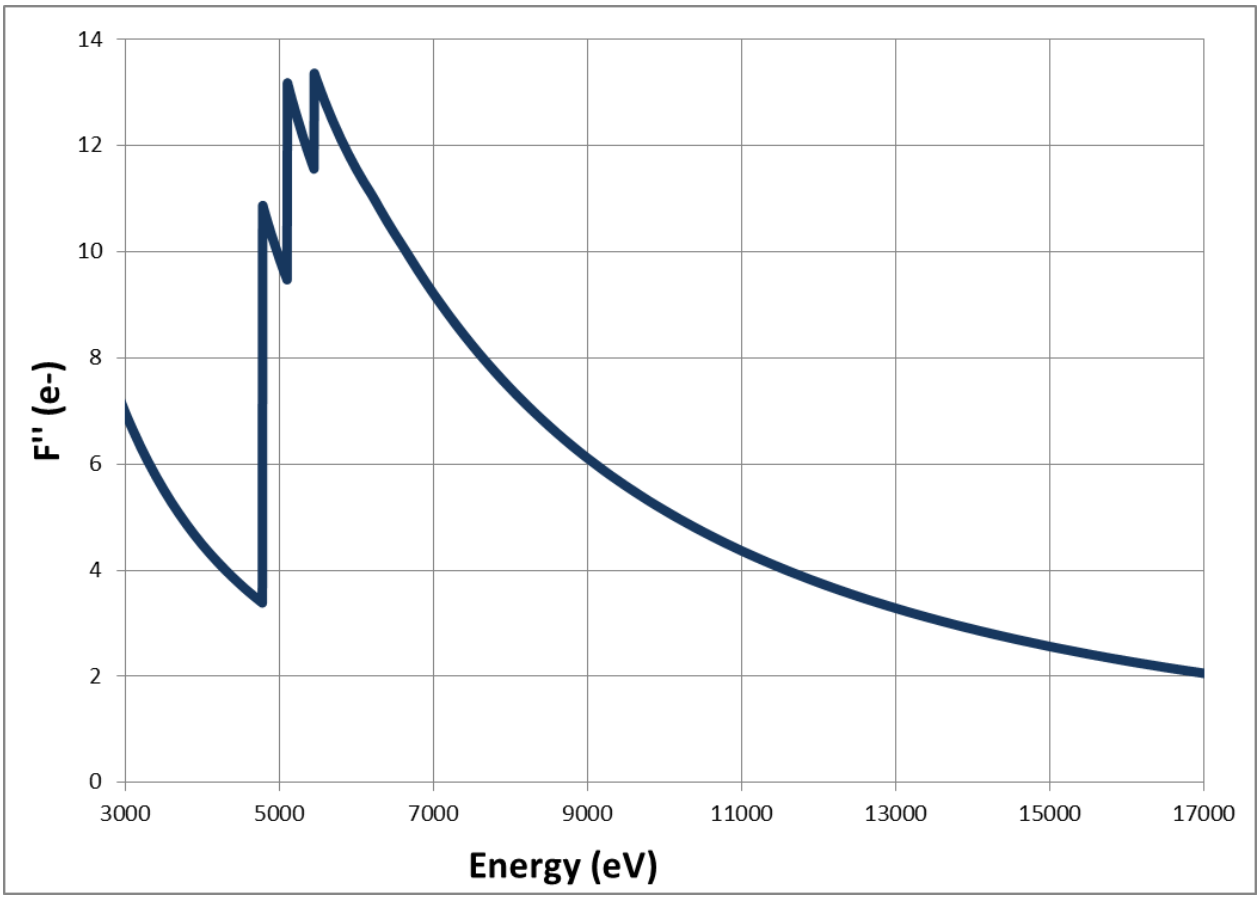


Figure 2.7 Theoretical anomalous signal for xenon

The anomalous signal coming from xenon increases as the x-ray energy is decreased and the L absorption edges are approached. A maximum value of f''=13.4e is reached at an x-ray energy of 5.45keV. Theoretically, data should be collected at an energy slightly above this to maximize anomalous signal. Due to absorption of scattered x-rays by air at such low x-ray energies, a higher energy must be chosen to compromise between x-ray absorption and anomalous signal from xenon.

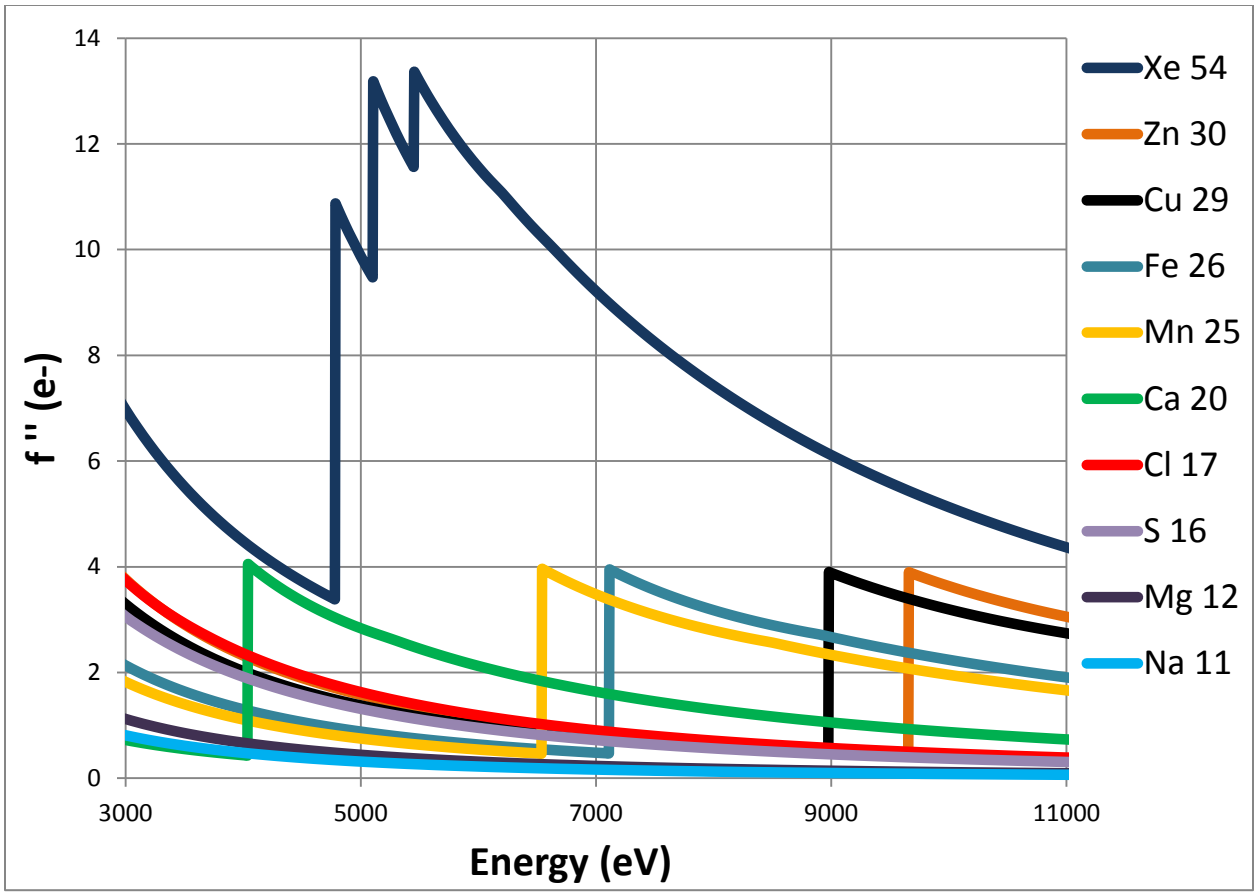


Figure 2.8 Theoretical anomalous signal for xenon and common atoms

There are many naturally occurring atoms in protein crystals, including atoms which are added during purification and crystallization, which contribute to anomalous signal at 7keV. Low occupancy xenon binding can produce an anomalous signal similar to these other atoms. In particular, very low occupancy xenon can become confused with chloride ions which show a preference for binding to the surface of proteins held in place by hydrophobic van der Waals or polar interactions[49].

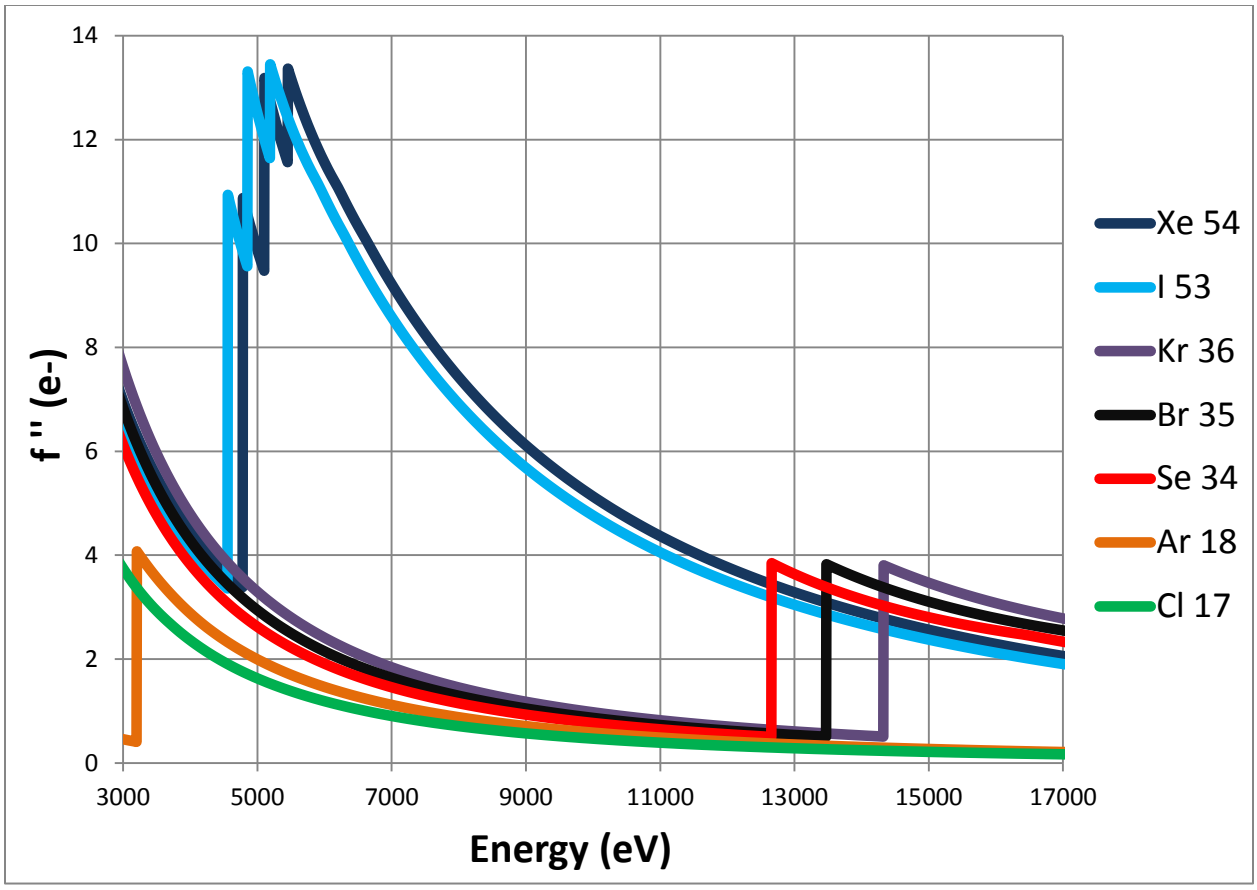


Figure 2.9 Theoretical anomalous signal for noble gases and halides

Xenon and iodine produce sufficient anomalous signal from a small number of partially occupied sites. Lighter group elements such as krypton and bromine have also been used successfully as sources of anomalous signal and are comparable in strength to selenium. Argon, although similar in strength to chlorine, has not been nearly as successful. In addition to a potentially small number of binding sites, its small size and polarizability make binding, based on induced and London interactions, extremely weak.

2.3.1 Xenon Pressurization

Derivatization by xenon gas has recently gained popularity in solving structures due to improvements in techniques which plagued early experiments. Pressurization chambers such as the Hampton Research Xenon Chamber mean no longer having to expose crystals to extreme changes in temperature and pressure and allows for much better measurement of gas pressure; in the older quartz capillaries, experimenters reported a failure rate of 70%[28,50]. One of two other major problems was that xenon gas within the capillaries

absorbed x-rays which meant long exposure times[25]. The other problem was the formation of xenon hydrate which is formed by water and xenon in closed cavities[51]. Both problems are solved using the Hampton Research Xenon Chamber in combination with modern cryocooling techniques whereby crystals are first pressurized by xenon gas and then quickly plunged into liquid nitrogen so as to "freeze in" the xenon.

Presently, SAD data has been used to solve nearly half of the structures determined by experimental phasing techniques[52]. Further, it has been estimated that approximately 50% of all proteins are able to bind noble gases based on studies performed on nine proteins of which five produced sufficient phasing information, two gave no phasing information indicating little or no xenon binding, and two were non-isomorphous[32]. For proteins which are difficult to phase, noble gases such as xenon are a reasonable method to be used to solve the phase problem.

3.0 Materials & Methods

Table 3.1 Proteins & reagents

Item	Supplier
100% Glycerol	Hampton Research
Ammonium Sulphate	Acros Organics
Ethylene Glycol	Hampton Research
Glucose Isomerase	Hampton Research
HEPES	Sigma-Aldrich
HEWL	Sigma-Aldrich
Potassium Sodium Tartrate	Sigma-Aldrich
Sodium Acetate Trihydrate	EMD Millipore
Sodium Chloride	EMD Millipore
Thaumatin	Sigma-Aldrich
Thermolysin	Hampton Research
Tris Base	Fisher Scientific

3.1 Proteins

3.1.1 HEWL

Hen egg white lysozyme is an enzyme which damages bacteria by attacking the carbohydrate chains which make up the bacterial cell wall. Breaking these carbohydrate chains causes bacteria to burst under their own internal pressure. High purity lysozyme powder can be purchased from commercial companies making it cheap and easy to grow crystals. Crystals grown from HEWL are particularly stable and durable for protein crystals and as such are ideal for teaching, testing, and demonstrations. For these reasons, lysozyme was selected as a protein standard on CMCF beamlines and as a test crystal for this study.

Lyophilized hen egg white lysozyme powder (product number HR7-110) was purchased from Hampton Research and used without further purification. Crystals were grown by the hanging drop vapour diffusion method (Figure 3.1) by mixing 6.5 µl reservoir solution containing 150 µl of NaCl (20% w/v), 50 µl of 1.0M sodium acetate (pH4.8), 125 µl of ethylene glycol and 175 µl water with 3.5 µl of 75 mg/ml protein solution in 0.1M sodium acetate (pH 4.8) (Crystallization protocol taken from Hampton Research website).

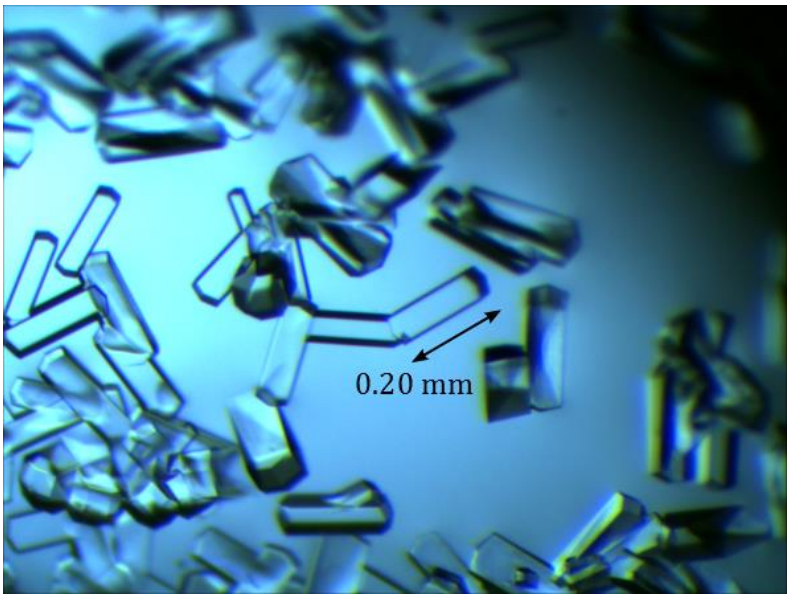


Figure 3.1 HEWL crystals

Crystals appeared within a couple of hours and grew to dimensions of approximately 0.20x0.10x0.10 mm^3 after 3 or 4 days.

3.1.2 Thermolysin

Thermolysin is a metalloproteinase enzyme containing one zinc ion and four calcium ions whose function is to catalyze the hydrolysis of peptide bonds of hydrophobic amino acids on the amino side. Thermolysin powder (product number HR7-098) was purchased from Hampton Research and used without further purification. Crystals were grown by the hanging drop vapour diffusion method (Figure 3.2) by mixing 4.6 µl of reservoir solution containing 187 µl of 2.0M ammonium sulphate, 50 µl of 1.0M tris (pH8.5), 120 µl of glycerol (50% v/v), and 143 µl of water with 5.4 µl of 25 mg/ml protein solution in 0.1M NaOH (Crystallization protocol taken from Hampton Research website).

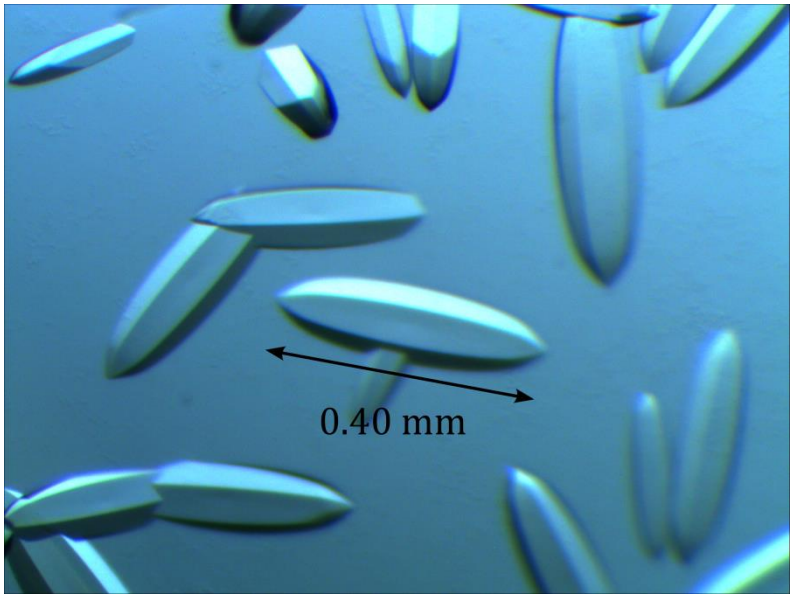


Figure 3.2 Thermolysin crystals

Crystals appeared within 24 hours and grew to dimensions of approximately 0.40x0.10x0.10 mm^3 after 3 or 4 days.

3.1.3 Glucose Isomerase

Glucose isomerase catalyzes the isomerization reaction between glucose and fructose. The reaction is reversible and goes to equilibrium so that there are approximately equal concentrations of glucose and fructose. Glucose isomerase can be crystallized by many different methods and by a wide variety of screen conditions which makes it another great candidate for study and learning. Glucose isomerase (product number HR7-102) was purchased from Hampton Research pre-crystallized (Figure 3.3) and supplied as a suspension in the following medium: 33 mg/ml in 6mM tris hydrochloride (pH adjusted to7.0), 0.91M ammonium sulphate, 1mM magnesium sulphate. Cryoprotection for glucose isomerase crystals was a combination of the supplied suspension medium mixed with glycerol for a glycerol concentration of 25% v/v.

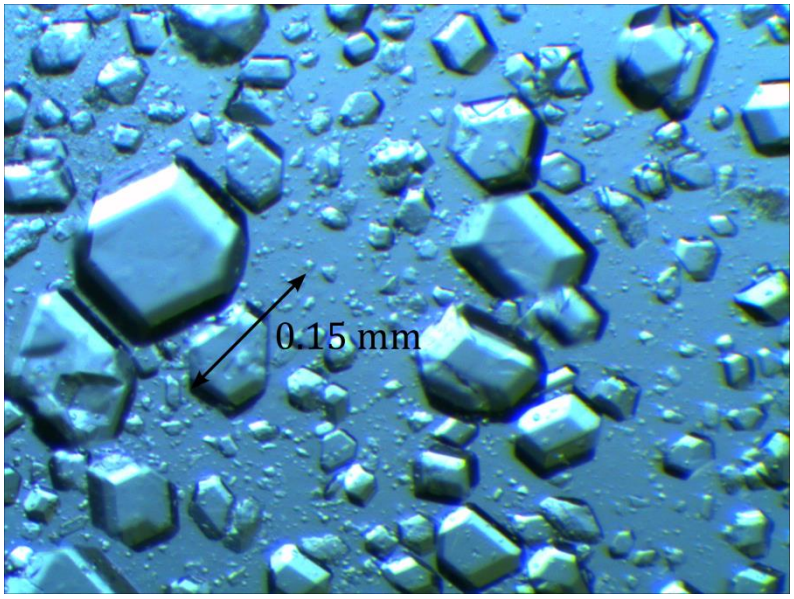


Figure 3.3 Glucose isomerase crystals

Crystals came pre-crystallized from Hampton Research. Suitable crystals selected for diffraction studies had dimensions of 0.15x0.15x0.15 mm^3.

3.1.4 Thaumatin II

The thaumatins are a family of proteins which are known mainly for their use as low calorie sweeteners and flavour modifiers. They are intensely sweet and although they have a different taste from sugar, they may be up to thousands of times sweeter. Thaumatin is primarily a combination of the proteins thaumatin I and thaumatin II the latter of which was used for this study. Thaumatin II powder (product number T7638-25mg) was purchased from Sigma Aldrich and used without further purification. Crystals were grown by the hanging drop vapour diffusion method (Figure 3.4) by mixing 5 µl of reservoir solution containing 50 µl of 0.5M Hepes (pH 7.0), 450 µl of 1.0M sodium potassium tartrate with 5 µl of 25 mg/ml protein solution in 0.5M Hepes (pH 7.0)[53]. Cryoprotection solution for thaumatin was composed of 130 µl of glycerol, 50 µl of 0.5M Hepes, and 320 µl of 1.0M sodium potassium tartrate (pH 7.0) for a final glycerol concentration of ~25% v/v.

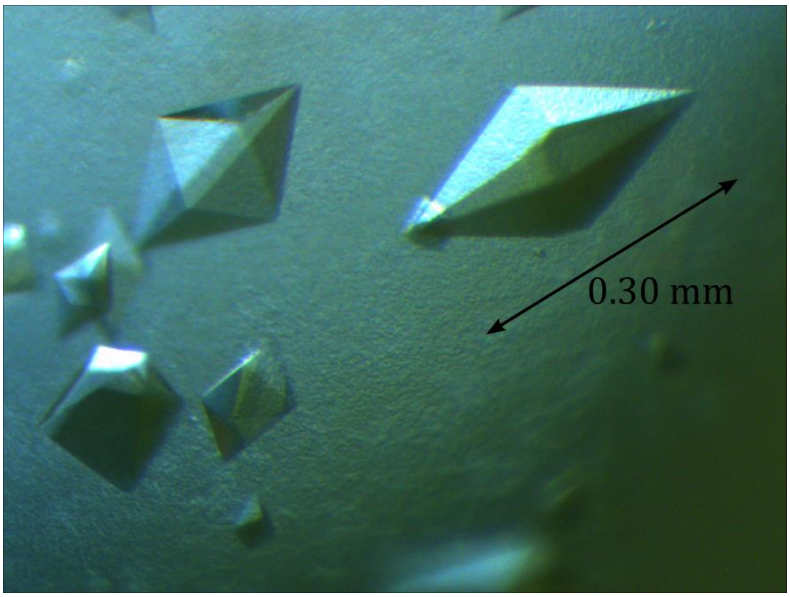


Figure 3.4 Thaumatin II crystals

Crystals appeared within 3 days and grew to dimensions of approximately 0.30x0.15x0.15 mm^3 after a week.

3.2 Xenon Gas

Research grade xenon gas was purchased from Praxair (product number RS-ELB50). This gas was used in combination with the Hampton Research Xenon Chamber and Xenon Recovery System to produce derivatives of each protein.

3.3 Hampton Research Xenon Chamber & Xenon Recovery System

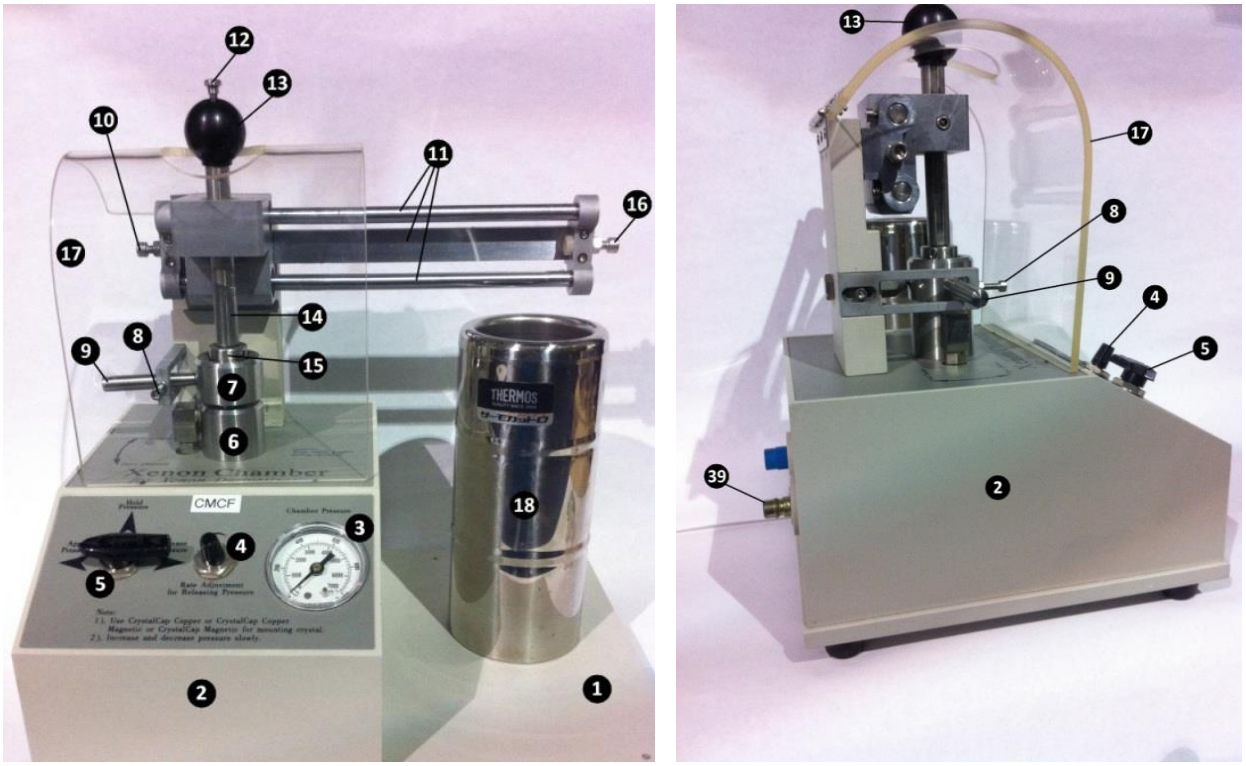
The Hampton Research Xenon Chamber™ (product number HR4-791) (Figure 3.5) is a pressure chamber designed to produce derivatives of protein crystals using xenon or other noble gases. The chamber is safely able to handle gas pressures up to 600psi above which the gas is vented via safety valve. The Xenon Recovery System™ (product number HR4-797) (Figure 3.6) was used in conjunction with the xenon chamber. The recovery system is a simple mechanical hand pump which helps to get the most out of xenon gas cylinders. The full setup and user manuals for the xenon chamber, xenon recovery system, and pressure regulator (product number HR4-793) (Figure 3.7) may be located and downloaded at the Hampton Research website.

Prior to Derivatization, setup and pre-operation steps were taken to ensure proper operation during derivatization. The xenon pressure regulator (Figure 3.7) was attached to the xenon tank using a crescent wrench. The other end of the pressure regulator where the black needle valve is located (32) was attached to the xenon recovery system inlet (26) by quick connector. The gas outlet on the xenon recovery system (22) was attached to the inlet on the xenon chamber (Figure 3.5, 39) also using a quick connector. At this point all valves on all equipment as well as the piston in the xenon recovery system were in the closed positions. The xenon chamber itself was left open with the magnetic base (15) moved out of the way along the glide tracks (11).

Next the valve on the xenon gas tank was opened followed by the pressure regulator T-valve (31) on the pressure regulator. This was indicated by the gas pressure on the xenon gas tank matching the tank gauge on the pressure regulator (29). The gas lines underwent a couple of purging cycles to ensure that there was only xenon in the lines. This process should be done sparingly since xenon gas is relatively costly and should not be wasted. Purging the gas lines was performed first by turning the 3 way valve on the xenon chamber (5) to "Apply Pressure to Chamber". The black needle valve on the pressure regulator (32) was opened and the 3 way valve on the recovery system (19) was set to "Fill Gas to Pump". This allowed gas to flow from the xenon tank through the pressure regulator and into the xenon recovery system. Turning the 3 way valve on the recovery (19) system to "Apply Gas to Chamber" cut off the gas from the xenon tank and allowed the gas in the recovery system to flow through the xenon chamber and out into the atmosphere. Repeating this process one or two more times effectively purged the gas lines.

Before harvesting crystals for derivatization, steps were taken to ensure that the crystals would not dehydrate and become damaged during derivatization. This was done simply by pipetting a small amount of mother liquor from the hanging drop vapour diffusion well into a small vial with wick paper (33) and placing it in the xenon chamber. To begin, a small square of wick paper was cut and placed around the interior of the xenon chamber vial. 200µL of mother liquor was pipetted into the vial and soaked up by the wick paper. Mother liquor was then added in 50µL increments and the vial turned upside down until the amount of mother liquor was almost able to form a drop and fall out of the vial. This ensured that there was sufficient mother liquor to provide humidity during derivatization

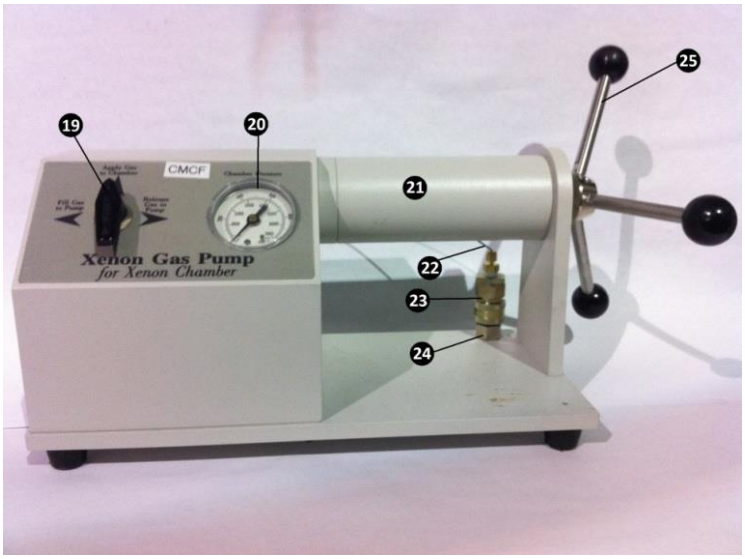
but not so much that the crystal would be sitting in liquid. The vial with wick paper and mother liquor was then placed inside the Xenon Chamber and the system is ready to produce derivatives.



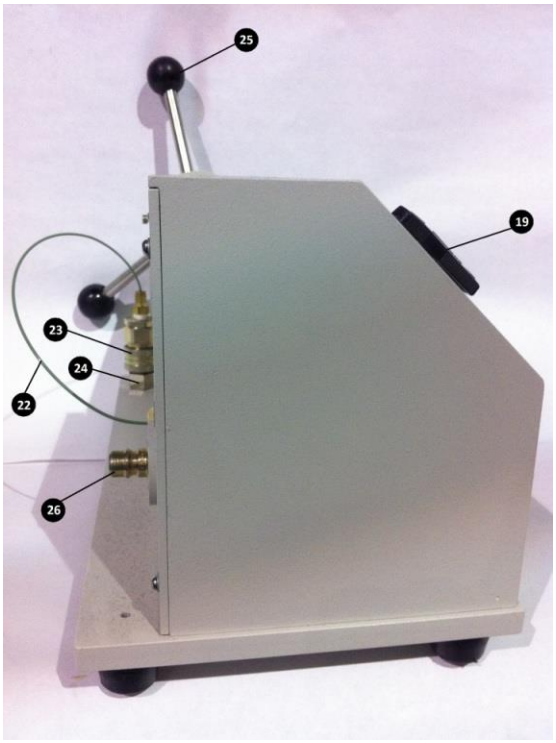
(a) Front view (b) Side View

Figure 3.5 Hampton Research Xenon Chamber

1. Dewar Dock	2. Chamber Platform	3. Chamber Pressure Gauge
4. Slow Release Gas Valve	5. 3 Way Valve	6. Xenon Chamber
7. Chamber Lock	8. Lock Lever Stop Screw	9. Lock Lever
10. Left Guide Stop Screw	11. Guide Tracks	12. Magnetic Base Release
13. Plunger Handle	14. Plunger	15. Magnetic Base
16. Right Guide Stop Screw	17. Safety Shield	18. Dewar
19. Dewar	39. Gas Inlet	



(a) Front view



(b) Side view

Figure 3.6 Hampton Research Xenon Recovery System

19. 3 Way Valve	20. Pressure Gauge	21. Piston Shell	22. Feed Line
23. Chamber Connector	24. Connector Base	25. Piston Handle	26. Gas Tank Inlet

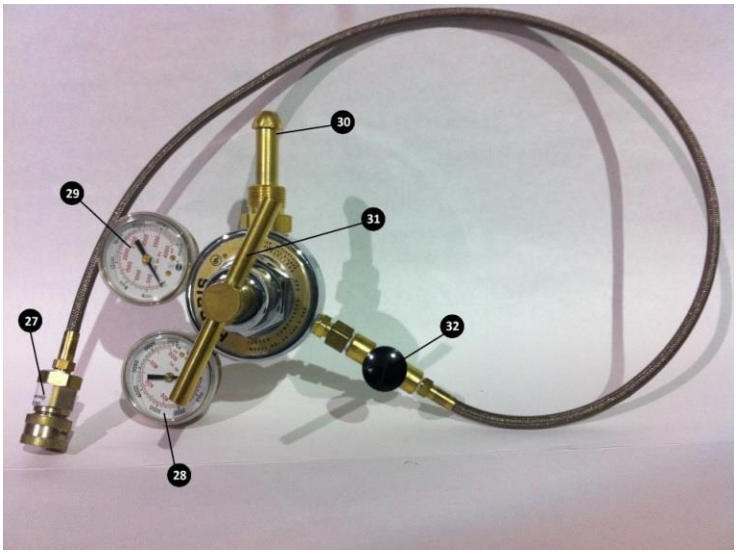


Figure 3.7 Xenon Gas Pressure Regulator

27. Quick Connector 28. Regulated Pressure Gauge 29. Xenon Tank Gauge
30. Xenon Gas Tank Inlet 31. Pressure Regulator Valve – T-Valve
32. Black Needle Valve

3.4 Crystal Harvest & Cryoprotection

Immediately prior to xenon derivatization, crystals were harvested using an 18mm Mounted CryoLoop™ of an appropriate nylon loop size such that the crystal roughly fills the entire loop. The CrystalCap Copper Magnetic™ with attached Mounted CryoLoop™ (34) was placed on the end of a magnetic wand (37) for easy manipulation during harvest. Crystals were harvested by hand using the magnetic wand under a microscope. Cryoprotection, if required, was performed prior to xenon derivatization since any attempt to cryoprotect crystals after xenon derivatization results in xenon gas diffusing out of crystals before they can be frozen.

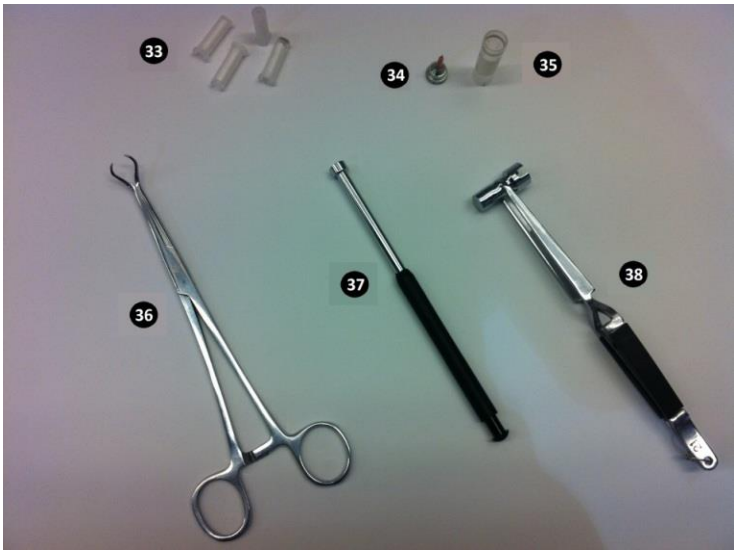


Figure 3.8 Cryocrystallography tools

33. Mini-Vial with wick paper
34. Crystal Cap Copper Magnetic™
35. Crystal Cap Copper Magnetic™ Vial with Magnetic Base
36. Crystal Cap Copper Magnetic™ Vial Clamp
37. Crystal Cap Copper Magnetic™ Wand
38. Crystal Cap Copper Magnetic™ Transfer Tool

3.5 Derivatization

Following harvest, CrystalCap Copper Magnetics™ with mounted crystals were removed from the magnetic wand by hand and placed on the magnetic base of the plunger (15) on the Xenon Chamber. The plunger (15) was slid along the guide tracks (11) to the left guide stop screw and lowered into the chamber. Once the chamber lock was closed (9), the sealed chamber was prepared to accept xenon gas. In order to get the most out of xenon gas tanks, the Xenon Chamber (Figure 3.5) was used in combination with the Xenon Recover System (Figure 3.6).

To begin, the piston handle on the Xenon Recovery System (25) was turned so that the piston was in the fully extended position. Then with the xenon gas tank valve, the pressure regulator T-valve (31), and the black needle valve (32) all set to open, the 3 point valve on the Xenon Recovery System (19) was set to the "Fill Gas to Pump". This allowed gas to flow freely from the xenon gas tank into the Xenon Recover System. When the desired gas

pressure had been reached as indicated by the pressure gauge on the Xenon Recovery System (20), the 3 point valve on the Xenon Chamber (5) was set to “Hold Pressure” and the 3 point valve on the Xenon Recovery System (19) was set to “Fill Gas to Pump”. This configuration trapped the xenon gas in the Xenon Recovery System neither allowing gas to flow from the xenon gas tank into the Xenon Recovery System nor out of the Xenon Recovery System into the Xenon Chamber. At this point, the gas pressure in the Xenon Recovery System did not need to be as high as the desired pressure for derivatization. The pressure could now be increased by turning the piston handle (25) without having to use additional xenon gas.

To begin derivatization, the 3 point valve on the Xenon Recovery System (19) should be set to “Fill Gas to Pump” followed by very slowly turning the 3 point valve on the Xenon Chamber (5) from “Hold Pressure” to “Apply Pressure”. The rush of xenon gas into the chamber caused by abrupt changes in gas pressure sometimes resulted in crystals being blown out of the nylon loop. A very slow change in pressure over several seconds was necessary to ensure that crystals were not lost and xenon gas was not wasted.

The pressure within the xenon chamber was monitored using the Xenon Chamber pressure gauge (3). To increase this pressure, the piston handle on the Xenon Recovery System (25) was turned by hand until the Xenon Chamber pressure gauge (3) reached the desired value for derivatization. The Xenon Chamber pressure was held constant by turning the 3 point valve on the Xenon Chamber (5) to “Hold Pressure” for the duration of derivatization.

Following derivatization, the xenon gas pressure was vented by turning the 3 way valve on the Xenon Chamber (5) to “Release Pressure”. This was also done very slowly over several seconds to prevent crystals from being blown out of the nylon loop. After the gas in the chamber had been completely vented, the chamber was opened using the lock lever (9). The plunger (14) was then raised and the CrystalCap Copper Magnetic™ (34) with mounted crystal was removed by hand, placed on the end of a magnetic wand (37), and quickly immersed in a nearby dish of liquid nitrogen to trap the xenon gas within the crystal for immediate data collection or for long term storage. Total time elapsed in between xenon gas pressure release and flash cooling in liquid nitrogen was not more than 5-8 seconds.

3.6 Crystallographic Software

3.6.1 MxDC

Mx Data Collector (MxDC) is the in house developed, graphical user interface software used on the CMCF beamlines for data collection. It allows users control over various aspects of beamline operation such as beamline setup, sample management, fluorescence scans, screening, data collection, and data processing[4,54].

3.6.1.1 Data Processing

The in house developed, automated data processing pipeline, AUTOPROCESS was used at CMCF[54]. The main engine of AUTOPROCESS is XDS[55] but it also makes use of other well established software packages such as XSCALE, XDSSTAT, BEST (strategy), CCP4 (ctruncate, f2mtz), POINTLESS (space group determination), and LABELIT.DISTL (ice ring detection). This allows for fully automated experiment planning and data processing requiring only minimal human intervention. AUTOPROCESS is also fully integrated with MxDC such that datasets collected using MxDC can be processed using AUTOPROCESS directly from the MxDC GUI. Processing reports are then available via the web through MxLIVE[4,54]. AUTOPROCESS is currently only available at the CLS but it has been used successfully on some datasets from other labs.

3.6.2 PHENIX

Python-based Hierarchical ENvironment for Integrated Xtallography (PHENIX) is a highly automated, python-based graphical user interface developed by Paul Adams *et al.* which hosts a variety of tools for the analysis, validation, and manipulation of x-ray diffraction data[56]. Additionally, PHENIX includes fully automated structure solution capability using the AutoSol wizard for experimental phasing and the AutoBuild wizard for model building and improvement thorough iterative model building, density modification, and refinement.

3.6.2.1 HySS

The Hybrid Substructure Search (HySS) is a highly-automated module in the Phenix package which locates heavy atom substructures. It requires a reflection file for input as well as the number of expected sites and outputs the heavy atom coordinates along with a correlation score describing how well the solution fits the data. The search method uses a systematic multi-trial approach using direct space Patterson followed by reciprocal space Patterson, then dual-space direct methods, and finally automatic comparison of solutions with automatic termination detection[56,57].

The method employed above begins with a test substructure. The Patterson map is calculated for the test substructure and is compared to the Patterson map computed from the data. Similarly, this type of calculation is performed in direct space using reflection intensities. Again a test substructure is created and its reflection intensities are calculated and compared to the data. The next step uses starting phases obtained from Patterson maps to obtain better than random initial phases for dual space recycling. To understand dual space recycling requires some knowledge of convolutions insofar as a convolution in real space is essentially multiplication in Fourier space. With that in mind, the Patterson function is a convolution in direct space that leads to squaring in reciprocal space as in equation 2.14. Now the opposite of that is the tangent formula (equation 2.13) where convolution in reciprocal space leads to squaring in direct space. This self-consistency leads to dual-space recycling where manipulation of phases in reciprocal space based on the tangent formula is recycled back and forth with direct space analysis[56,57].

For actual substructure construction, a two-atom substructure is chosen at random knowing only the positions of the two atoms relative to one another within the unit cell. The substructure is then translated randomly throughout the unit cell and is given a correlation score for each position. Using a pair of two-atom substructure coordinates which scored well, the previous step is repeated for a third atom by adding a third atom to the substructure while keeping the first two atoms fixed. This process is repeated until the desired number of atoms in the substructure has been generated[56,57].

Substructures which score well are selected for dual space recycling. Dual space recycling corrects inaccuracies in the initial substructure and itself is graded as a correlation between calculated and observed intensities. Finally, HySS employs the random omit

technique where substructure atoms are randomly omitted by only searching for 90% of the desired number of atoms of which 2/3 are randomly selected for recycling[56,57].

3.6.2.2 AutoSol

The AutoSol wizard is a highly automated experimental phasing module which allows much flexibility for advanced users. AutoSol requires only a reflection file to run but other files such as sequence files and heavy atom coordinate .pdb files may also be provided for improved performance. AutoSol then runs a series of PHENIX tools beginning with phenix.solve to scale data sets, followed by phenix.xtriage for twinning analysis and anisotropy correction, and then phenix.hyss which determines the heavy atom substructure if not already provided by phenix.hyss. *Ab initio* phasing is then carried out by phenix.phaser or phenix.solve followed by density modification performed by phenix.resolve. Finally, a round of automated model building is performed by AutoBuild[56,58].

3.6.2.2.1 Figure of Merit

The method used by AutoSol for determining the success of initial phases is known as a figure of merit (FOM). The FOM is a numerical score from 0 to 1 and is defined as the cosine of the lack of closure error, $\varepsilon(\varphi)$, which results from the discrepancy between theoretical and observed values as shown in Figure 3.9. A FOM of 0.3 is near the minimum value for useful phase information and corresponds to $\varepsilon(\varphi)=72.5$. Figure of merit values greater than 0.3 are good and a figure of merit greater than 0.5, $\varepsilon(\varphi)=60$, is very good[56,59].

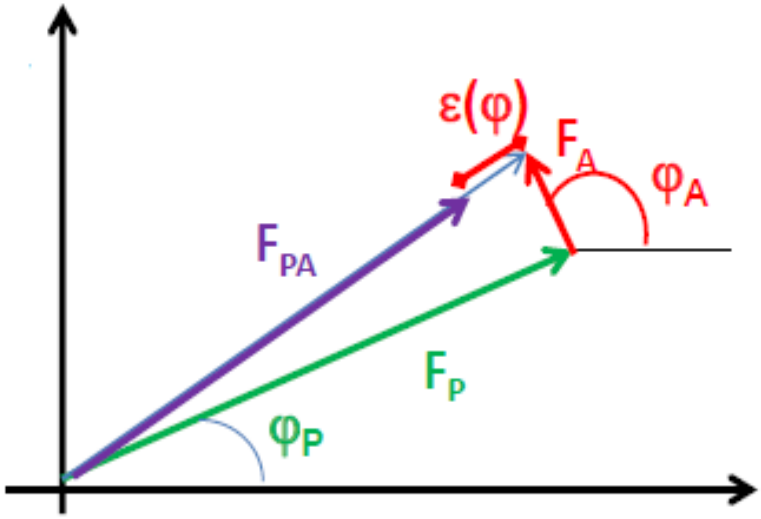


Figure 3.9 Lack of closure error[59]

A vector diagram representation of the lack of closure error, ε(φ). In theory, the vectors representing the structure factors for the protein, $\mathbf{F_P}$, and the anomalous substructure, $\mathbf{F_A}$, should add vectorially to give the structure factors for the protein and anomalous signal together, $\mathbf{F_{PA}}$. Experimentally these vector quantities do not add perfectly and ε(φ) is a measure of how well observation agrees with theory.

3.6.2.3 AutoBuild

The AutoBuild wizard is a model building tool in the PHENIX package which uses phenix.resolve for iterative model building and statistical density modification with phenix.refine for refinement. AutoBuild requires the experimental data, an initial electron density map, and starting model to run; most of these files are piped directly into AutoBuild from AutoSol output files. Additionally, files such as amino acid sequence files or .pdb files of partial models may be provided for improved performance[56,60].

3.6.2.4 phenix.refine

A standard settings run of phenix.refine is called automatically by AutoBuild as part of the iterative model building process. Phenix.refine may also be run separately for additional features such as in refinement of f' and f'' values for anomalously scattering atoms. To run separately, phenix.refine requires only an experimental data file which includes Rfree

values and phases as well as a file containing the starting model coordinates. Both of these files may be piped directly into phenix.refine from AutoBuild output files. While run separately, each round of phenix.refine displays the updated model in accompanied molecular graphics software such as Coot[61] or PyMol[56,62].

3.6.3 Coot

Coot is a molecular graphics program used in macromolecular model building and validation. The software displays electron density maps and atomic models while providing a variety of manual and automated model building tools. Coot will also interface with other programs such as phenix.refine for refinement, validation, and graphics[61].

3.6.4 PyMol

PyMol is molecular visualization software used for production of publication quality images of molecular models[62].

4.0 Results

4.1 Optimal Conditions for Xenon Derivatization

Xenon binding is essentially complete on a time scale of minutes and therefore should not require xenon pressurization for longer than 5 minutes[18,33–35]. Additionally, derivatives had previously been produced using a wide range of xenon pressures (1-100 bar or 14.5-1450 psi)[18]. Data were collected for a total of 33 HEWL crystals with xenon gas pressures ranging from 50-500 psi with full data tables located in Appendix A. It was discovered that subjecting crystals to high pressure (500 psi) for a long period of time (60 min) neither improved nor adversely affected the crystal or data quality. Further, 150 psi was determined to be the minimum pressure required to successfully phase HEWL. Xenon pressures below 150 psi yielded insufficient anomalous signal to solve the phase problem. Therefore it was decided that 300 psi was the optimal pressure for xenon phasing without subjecting crystals to unnecessarily high gas pressures.

4.2 Data Collection

Diffraction data were collected at the Canadian Macromolecular Crystallography Facility (CMCF) at the Canadian Light Source (CLS) on the 08B1-1 (CMCF-BM) beamline. Mounted crystals were attached to the magnetic goniometer head using a Crystal Cap Copper Magnetic™ Transfer Tool (38) and were maintained at 100K for the duration of data collection using an Oxford Instruments Cryojet. Data collection parameters and automated data collection was handled using Mx Data Collector (MxDC). Data collection strategy was determined by AUTOPROCESS using six images that were taken in 0.5° pairs at 45° apart instead of consecutively; this method gives a better sampling of the crystal lattice. A complete table of data collection parameters and processing results can be located in Appendix A.

4.3 Data Processing

Initial data processing, including generation of reflection (.mtz) files, was performed using AUTOPROCESS. Diffraction data were collected for a total of 45 protein crystals of which

33 were HEWL; HEWL was used mainly as a test protein in order to determine optimal data collection and processing parameters. Initial data processing was performed on all 45 crystals, including determination of number and location of xenon sites, using HySS. Using XDS[55], the strength of the anomalous signal, SigAno, for each resolution shell was calculated; this was the first indicator of potential xenon binding. Multiple runs of HySS were performed for each protein looking for anywhere between 2 to 25 xenon sites. The xenon coordinates which had the highest correlation with the data and the fewest number of sites were selected for input into AutoSol.

Input files provided for AutoSol included the experimental reflection data (.mtz) and the xenon site coordinates (.pdb) determined using HySS. A text file containing the one letter amino acid sequence for the protein, known as a sequence file (.seq), was included for all runs of AutoSol; a sequence file was not necessary for AutoSol to run however since all proteins were previously known, a sequence file allowed AutoSol to run automated model building more quickly. In the case of determining protein structures *de novo*, AutoSol would take longer as would manual rounds of model building and refinement.

Phasing scores based on the xenon sites determined using HySS were determined by AutoSol as a figure of merit (FOM); these FOMs are a score of how well the xenon coordinates solve the phase problem. This score is the second indicator of whether or not there are anomalous scattering atoms in the protein. For each type of protein studied, the data set which produced the highest FOM was selected for further model building and refinement. A complete data table including FOM for each protein crystal can be located in Appendix B.

4.4 Model Building & Refinement

Automated model building and refinement was performed by AutoBuild. Input files for AutoBuild included the experimental data (.mtz), the initial electron density map (.mtz), and the starting model coordinates including heavy atom coordinates (.pdb). A sequence file (.seq) was once again provided. A standard run of AutoBuild consisting of 3 refinement cycles, 6 maximum iterative build cycles, and 15 maximum rebuild cycles was performed on one dataset for each protein. All water molecules added by AutoBuild, as well as any

heavy atom coordinates which were determined to have zero occupancy by either AutoSol or AutoBuild, were removed from the models and these working models were then inspected using Coot.

Protein models still had missing or incorrect amino acids after automated model building particularly those which were terminal residues or residues which coordinate some metals. For manual model improvement, coordinates were downloaded from the PDB for each protein and these models were superimposed on the working models using Coot's LSQ (least squares fit) feature. The number of models which was available for comparison for each of the proteins depended on previous work performed by other researchers and sometimes even included xenon derivatives for direct comparison. The refined experimental electron density and the 2Fo-Fc maps were used to determine the correct positions and conformations of amino acids. In most cases, models from the PDB agreed both with one another as well as with the experimental data. When PDB models did not agree with experimental data, the working model was built to fit the experimental data as well as possible although PDB models still served as a guide in model building.

The models were improved to fit the data in Coot as well as possible. Manual fitting was followed by a run of phenix.refine for 8 refinement cycles; additional cycles showed little to no improvement and 8 cycles was chosen for all subsequent runs of phenix.refine. Alternation between manual fitting in Coot and automated refinement using phenix.refine was performed until all non-solvent peaks in the 2Fo-Fc were explained by the working model and additional runs of phenix.refine showed over fitting or little to no improvement. If models from the PDB showed solvent molecules other than waters, then these molecules were only added to the working model when the molecule was present in the crystallization conditions and it was obvious that the solvent molecule fit the experimental data. In most cases, it was not obvious that these solvent molecules were present in the experimental data and these regions were described instead by waters. A final run of phenix.refine was performed to automatically add waters back into the model. One last inspection of the model in Coot was performed and any waters which did not agree with the 2Fo-Fc maps were added or removed as needed. Estimation of the weakest signal which could still be meaningfully described by water molecules was approximated by randomly omitting several more obvious waters. The peaks in the 2Fo-Fc map which

described these missing waters ranged from 8 to 15 rmsd and therefore it was assumed that anything below 5 rmsd was either very low occupancy water or else map noise. Model building and refinement was complete after all peaks in the 2Fo-Fc map above 5 rmsd were explained.

4.4.1 Anomalous Refinement

A final round of refinement using phenix.refine was performed on each protein in order to differentiate between xenon sites and other anomalously scattering atom sites. Phenix.refine has the capability to refine the anomalous signal, f'', coming from anomalous scattering atoms and was run as described earlier for 8 refinement cycles but this time including refinement of anomalous groups. Each anomalously scattering atom site was defined as its own anomalous group so that each would be refined separately. An initial value of f''=0.0 was provided to phenix.refine; this value was chosen so that anomalous refinement would not be biased. Furthermore, it was not certain which sites nor how many were actually xenon and not some other anomalously scattering atom.

PHENIX is cutting edge crystrallographic software that is constantly being improved upon[56]. Over the course of learning to use it, the PHENIX package has become increasingly robust with many features and options added. Enabling alpha test mode in the PHENIX software allowed testing of the new features as was the case for the final step of refining anomalous groups, in particular. Refinement of anomalous groups was being worked on presently and mail correspondence between myself and PHENIX developers was necessary to work out some bugs in this feature. Full data collection statistics are shown in Table 4.1.

Table 4.1 Data collection statistics

	Lysozyme	Thermolysin	Glucose Isomerase	Thaumatin II
Data-collection parameters				
Beamline	08B1-1 (CMCF-BM), CLS	08B1-1 (CMCF-BM), CLS	08B1-1(CMCF-BM), CLS	08B1-1 (CMCF-BM), CLS
Energy (keV)	7	7	7	7
Wavelength (Å)	1.7712	1.7712	1.7712	1.7712
Temperature (K)	100	100	100	100
Oscillation range (°)	0.5	0.5	0.5	0.5
No. of images	540	542	551	580
Exposure time per frame (s)	4	4	8	3
Data-integration statistics				
Space group	$P4_32_12$	$P6_122$	I222	$P4_12_12$
Unit Cell parameters				
a (Å)	79.13	92.93	92.66	57.87
b (Å)	79.13	92.93	98.07	57.87
c (Å)	36.92	129.67	101.94	150.23
α (°)	90	90	90	90
β (°)	90	90	90	90
γ (°)	90	120	90	90
Resolution limits (Å)	39.57-2.02 (2.15-2.02)	46.45-2.03 (2.15-2.03)	48.98-2.03 (2.15-2.03)	45.83-2.03 (2.15-2.03)
Total no. of reflections	130936	552511	269427	308718
No. of unique reflections	12886	40015	53137	30059
Multiplicity	10.2	13.8	5.1	10.3
Completeness (%)	88.2 (51.5)	98.9 (93.2)	91.6 (56.7)	95.4 (73.9)
R_{meas}* (%)	3.1 (7.2)	5.1 (20.4)	7.8 (39.6)	2.7 (4.4)
Mean I/σ(I)	59.2 (30.4)	39.8 (10.2)	22.5 (4.4)	70.3 (39.2)
Molecules per asymmetric unit	1	1	1	1
Solvent content (%)	38.0	45.9	53.7	54.5
Refinement				
Resolution (Å)	39.57-2.02	46.45-2.03	48.98-2.03	45.83-2.03
Rwork/Rfree	0.14/0.22	0.13/0.18	0.13/0.18	0.12/0.17
No. of reflections	11580	36500	50520	27116
No. of atoms	1178	2900	3549	1937
B factor from Wilson plot ($Å^2$)	12.2	16.1	12.5	9.88
B factor ($Å^2$)	16.4	21.9	20.5	14.5
Estimated coordinate error (Å)	0.2	0.19	0.17	0.17
R.m.s.d. bond lengths (Å)	0.006	0.008	0.007	0.007
R.m.s.d. angles (°)	1.027	1.124	1.021	1.054
Ramachandran plot				
Favoured (%)	97.6	96.2	96.9	98.6
Allowed (%)	2.4	3.8	3.1	1.4
Xenon Bound	Yes	Yes	Yes	No

Bijovet pairs were not merged for data processing.
Cutoff based on experimental geometry constraints at 7keV.

5.0 Discussion

5.1 Proteins

HEWL was selected as the main test protein for several reasons. Learning how to grow protein crystals was the initial goal. Lysozyme crystalizes quickly producing robust crystals which are ideal for running tests. Secondly, a test protein provided the opportunity to establish a method and to hone data collection and processing techniques. Finally, lysozyme is known to bind xenon. The PDB contains a large number of deposited lysozyme structures due to its being such a common test protein. This includes a lysozyme-xenon derivative (1C10[30]) which shows three binding sites for xenon. Thermolysin was selected from a list of proteins which were known to bind xenon and of which only thermolysin could be crystallized. In addition to these, thaumatin II and pre-crystallized glucose isomerase were available for study although neither had previously been shown to bind xenon. In total, data were collected for 45 protein crystals of which 33 were lysozyme, 9 were thaumatin II, 2 were thermolysin, and 1 was glucose isomerase. A structure was built for each protein using the anomalous signal at 7keV.

5.2 Wavelength selection

As shown in Figure 2.7, xenon has a very high anomalous signal in the range of most crystallography beamlines. Xenon's high anomalous signal means strong anomalous data can be obtained even very far from any absorption edges. In xenon's case, the anomalous signal increases with increasing wavelength until the L-I edge located at 2.27Å. It was experimentally established that without the use of a helium box, the optimum energy for low energy data collection is 7keV[63] based on a compromise between anomalous signal from xenon and absorption of scattered x-rays.

Data collection at 7keV (1.77 Å) provided sufficient anomalous signal to solve the phase problem for each of the four proteins and it was not necessary to attempt data collection at a lower energy. Had phasing been unsuccessful, data collection could have been attempted at lower energies. The use of a helium box would reduce the absorption of scattered x-rays allowing data collection closer to the L-I edge. This would be particularly beneficial for

extremely low occupancy xenon sites or for very large proteins with a low number of xenon binding sites.

5.3 Anomalous signal

The first indicator of xenon binding was determined using XDS (Figure 5.1) as the mean anomalous difference (SigAno) in units of the estimated standard deviation, σ:

$$SigAno = \frac{|F(+)-F(-)|}{\sigma} \qquad (5.1)$$

SigAno is the total anomalous signal from all anomalous scatterers in the protein and therefore it is still not known whether or not xenon binding has occurred. A strong anomalous signal at this stage is important, however, as a model of the protein may be built which allows further analysis of anomalous substructure. For each of the four proteins examined, there was sufficient anomalous signal to move on to substructure determination.

Figure 5.1 Anomalous signal vs Resolution

AUTOPROCESS calculates anomalous signal and breaks it up into resolution bins. A graph summarizing the data shows that anomalous signal peaks in the mid-low resolution range. A sharp drop off in anomalous signal occurs around 3-2Å for most of the proteins illustrating the importance of obtaining good quality mid-low resolution data.

5.4 Anomalous substructure

For substructure determination, all anomalous scatterers were treated as xenon. It was not yet known which xenon sites were true partially occupied xenon sites and not simply some other anomalous scatterer whose anomalous signal resembles that of a partially occupied xenon site (Figure 2.8). Ideally when looking at a graph of the xenon occupancy vs site number, one would hope to see a small number of high occupancy sites followed by a steep drop; this may indicate xenon binding[18]. Next would come several smaller peaks due to other anomalous scatterers and possibly some medium to low occupancy xenon. Finally,

any very low occupancy sites would usually be noise which would be removed in subsequent model building and refinement.

Initial data (Figure 5.2) from test protein HEWL shows 1-3 possible binding sites for xenon followed by a number of lower occupancy sites. Some of these lower occupancy sites may be heavier atoms or solvent. Also, given that HEWL was the only protein in which chlorides were present during crystallization, it is likely that many of these low occupancy xenon sites are actually chloride.

Thermolysin and glucose isomerase proved to be much better test proteins as they both have one high occupancy site followed by a steep drop which suggests at least one binding site for xenon. The remaining few sites are likely a mix of lower occupancy xenon sites and metals. The curve of thaumatin II suggests little to no xenon binding even though thaumatin II had a strong anomalous signal previously. The strong anomalous signal in thaumatin II could be a combination of many low occupancy xenon sites although this is unlikely.

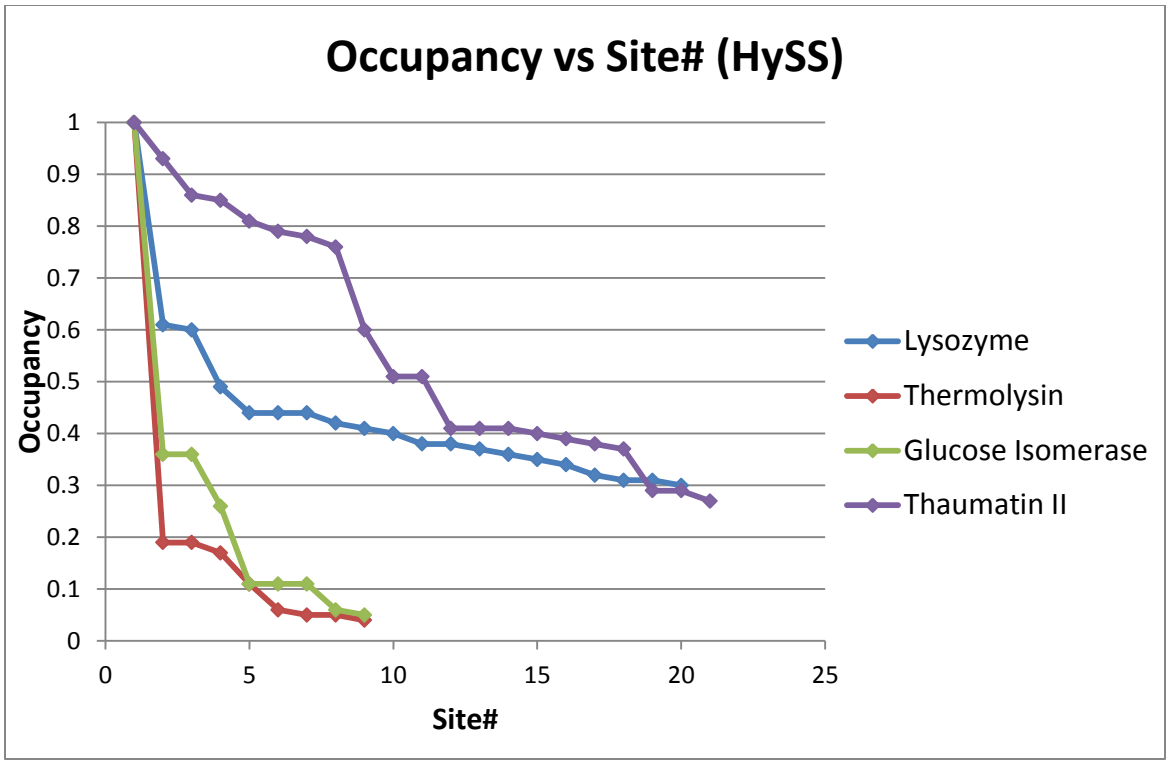


Figure 5.2 Xenon site occupancy vs Site number

HySS finds the xenon sites using anomalous maps which have been calculated from anomalous data. It treats all anomalous peaks as xenon and normalizes the scale by giving the largest off-origin peak an occupancy of 1 regardless of actual occupancy. All other peaks are then some fraction of the largest off-origin peak.

5.5 Figure of Merit

The anomalous substructures as determined by HySS yielded good FOMs for each of the four proteins (Table 5.7). The slightly lower FOM coming from the glucose isomerase was partially due to the lower multiplicity of the data collected and not necessarily due to the absence of anomalous scatterers. Upon solving the phase problem, initial models of each protein were built automatically using AutoSol and approximate starting occupancies for each anomalous scatterer were calculated. It is common during cycles of model building and refinement for Phenix software to append or remove anomalous scatterers from the anomalous substructure model in order to better fit the data[56,59].

Test protein HEWL showed a pair of very low occupancy sites followed by a number of lower occupancy sites (Figure 5.3). This indicates two potential xenon sites followed likely by a combination of chloride ions and native sulfurs. In the case of thaumatin II, there are no high occupancy peaks and because of that, as with the case of lysozyme, a large number of peaks belonging to native sulfurs are included in the anomalous substructure as they are providing the majority of the anomalous signal. Once again, thermolysin and glucose isomerase were very good test proteins each showing a single high occupancy peak indicative of xenon binding. The remaining peaks are likely a combination of lower occupancy xenon sites, metals, and native sulfurs.

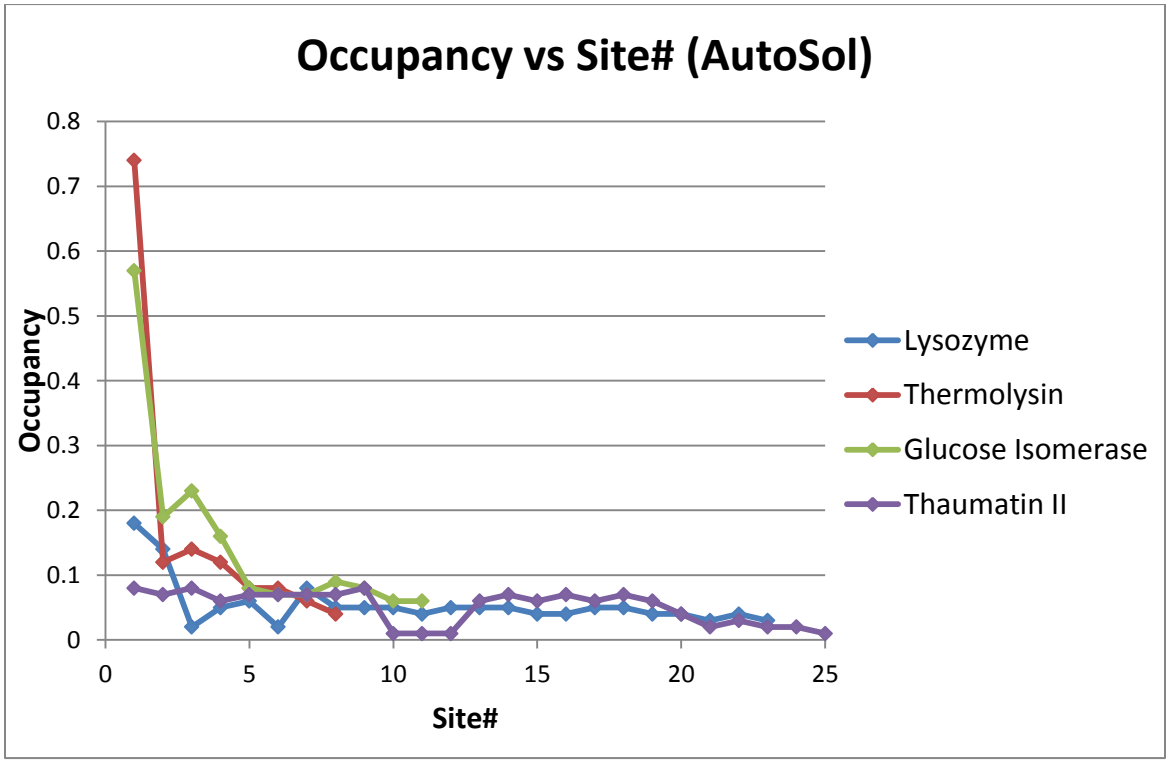


Figure 5.3 Xenon site occupancy vs Site number

AutoSol builds a model of the protein using the anomalous data and the anomalous atom sites determined by HySS. Once again, all anomalous scatterers are treated as xenon and each is given an approximate occupancy based on the initial model. The large number of sites in lysozyme and thaumatin II is due to the lack of any strong anomalous scatterers in the protein. This results in most of the anomalous signal being provided by small native sulfurs and, in the case of lysozyme, chlorides as well.

Refinement to convergence saw many sites removed from the substructure as they either belonged to native sulfurs or were determined to have zero occupancy[56]. Thermolysin was the best case as there is clearly a single, high-occupancy xenon site (Figure 5.4). The remaining thermolysin sites as well as all sites for the other three proteins were not clearly distinguishable as xenon and further investigation was needed. The large number of sites remaining for lysozyme is due to chlorides which were present in crystallization. In the case of thaumatin II, all of the low number sites were removed from the substructure having mostly been native sulfurs. The remaining thaumatin II sites, despite several decent occupancies, are most likely just noise.

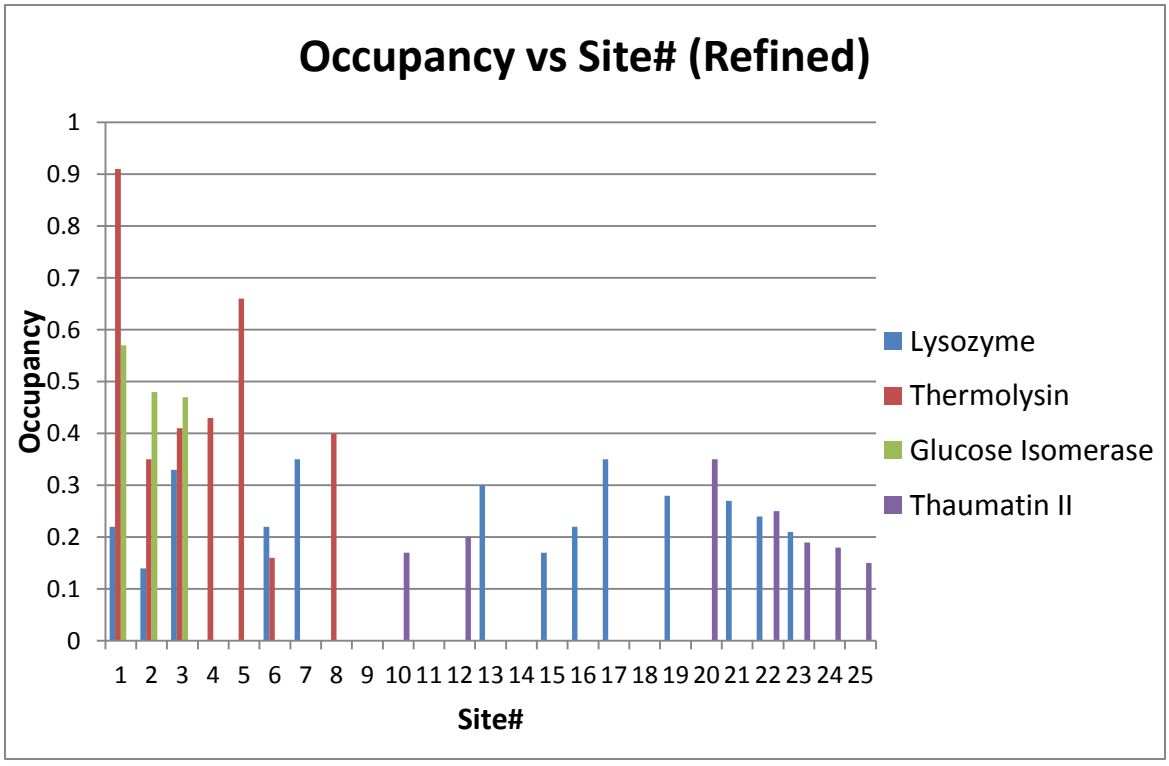


Figure 5.4 Xenon site occupancy vs Site number

Refinement to convergence removed many peaks leaving only those which were likely not noise.

5.6 Anomalous Refinement

The refinement of f'' values shed some light on the identities of the anomalous scatterers (Figure 5.5). The results for test protein HEWL were promising. For the first time, refinement of f'' values suggest strong likelihood that sites 1 and 2 are probably true, albeit very low occupancy, xenon sites when compared to the other remaining sites. Thaumatin II initially had a very strong anomalous signal followed by a good FOM. However, of an initial 25 sites, 17 turned out to be native sulfurs. The remaining 8 sites are low occupancy, low refined f'', and are likely not actual sites.

Once again, thermolysin site 1 has a very high f'' providing further evidence that this is almost certainly a true xenon site. The remaining thermolysin sites are most likely a combination of metals having relatively small refined f'' compared to site 1. Site 5 shows a high occupancy (Figure 5.4) with a low f'' which indicates that this is not a xenon site. Site 6 on the other hand is interesting because of its low occupancy (Figure 5.4) compared with its relatively high refined f'' value (Figure 5.5). This may indicate a second binding site since the software seems to have some difficulty when refining f'' values for low occupancy sites. In the case of glucose isomerase, neither the occupancies nor the refined f'' values helped to distinguish any of the remaining sites as xenon (Figures 5.4 & 5.5). There remains the possibility that none or all of the remaining glucose isomerase sites are xenons and thus further investigation is required.

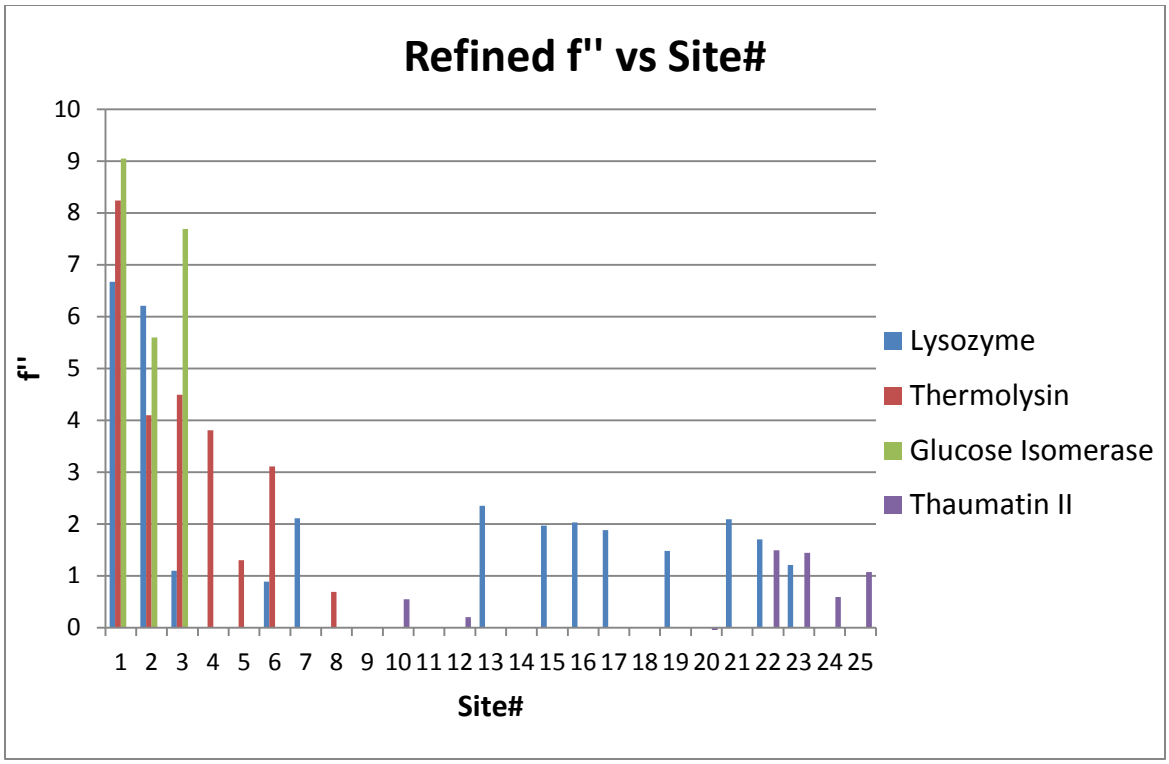


Figure 5.5 Refined f'' vs Site number

F'' refinement in the PHENIX package was still in alpha testing stages at the time of data processing. Although not perfectly accurate, these refined values were still useful when trying to distinguish xenon from other anomalous scatterers.

5.7 Binding site inspection

Table 5.1 Coordination geometry of common metals[64,65]

Element	Coordination Geometry	Ligand
Zn	4, 5, 6	S, N, O
Mg	6	O
Ca	7, 8, 9	O preferred
Cu	2, 4, 6	-
Mn	4, 6	-
Fe	6	-
Cl	van der Waals[49]	
Xe	van der Waals	

Different metals prefer different numbers and types of ligands. Xenon gas and chloride ions behave differently and prefer to bind based on weak van der Waals forces.

Table 5.1 shows the preferred coordination of common metals including xenon and chloride. Table 5.2 shows known protein-metal bond distances and to which atoms they bind. Xenon does not bond in this manner and instead interacts via van der Waals interactions based on the van der Waals radii where distances between xenon and protein carbon, oxygen, and nitrogen are the combined xenon-ligand radii of 3.86Å, 3.68Å, and 3.71Å, respectively. Based on the aforementioned criteria, binding sites which were likely metals were not included in further model building analysis. Except for a few obvious xenon binding sites, the remaining sites must be inspected visually in the refined models. Overall, observed protein-metal bond lengths were approximately 10-15% longer than expected bond lengths based on statistics taken from the PDB and CSD (Table 5.2).

Table 5.2 CSD and PDB bond length statistics for protein-metal complexes[66]

	O, water	O, Asp or Glu monodentate	O, main-chain carbonyl	N, histidine	S, cysteine
Na	2.41**	2.41**	2.38*	-	-
Mg	2.07***	2.07**	2.26	-	-
K	2.81*	2.82*	2.74*	-	-
Ca	2.39**	2.36**	2.36**	-	-
Mn	2.19***	2.15***	2.19	2.21**	2.35
Fe	2.09**	2.04**	2.04	2.16*	2.30**
Co	2.09**	2.05***	2.08	2.14**	2.25*
Cu	2.13	1.99*	2.04	2.02**	2.15
Zn	2.09***	1.99***	2.07	2.03***	2.31**

*** most reliable, std dev ~0.05Å
** std dev 0.10Å
* std dev 0.15-0.20Å
no marking is least reliable

Small revisions to predicted distances around metal sites in proteins

Test protein HEWL sites Xe1 and Xe2 are likely bound xenon from data summarized in Figures 5.3-5.5. The anomalous maps in Figures 5.6 and 5.8 show the presence of anomalous scatterers albeit very weak ones. The weak anomalous signal is the result of these both being very low occupancy xenon sites set apart from the other sites only by their relatively high refined F'' values(Figure 5.5). Xe1 (Figure 5.7) has an occupancy of 0.22 and is hydrophobic in nature. All of the nearest contacts are to carbon atoms and there are no contacts within 3.4Å. This distance is too far to be a metal-protein bond and there is no coordination geometry present.

HEWL site Xe2 (Figure 5.9) is very different from Xe1. The site is made up of two adjacent protein molecules which produce a pocket for xenon along a solvent channel. Two arginines and two threonines from neighbouring protein molecules produce a pair of beta sheets between which xenon can bind (Figure 5.10). This pocket, composed mostly of oxygens and nitrogens, shows no coordination geometry typical of metal binding and is

larger than Xe1 having no contacts within 3.9Å. The low 0.14 occupancy of this site may be the result of the site being slightly too large for xenon atoms.

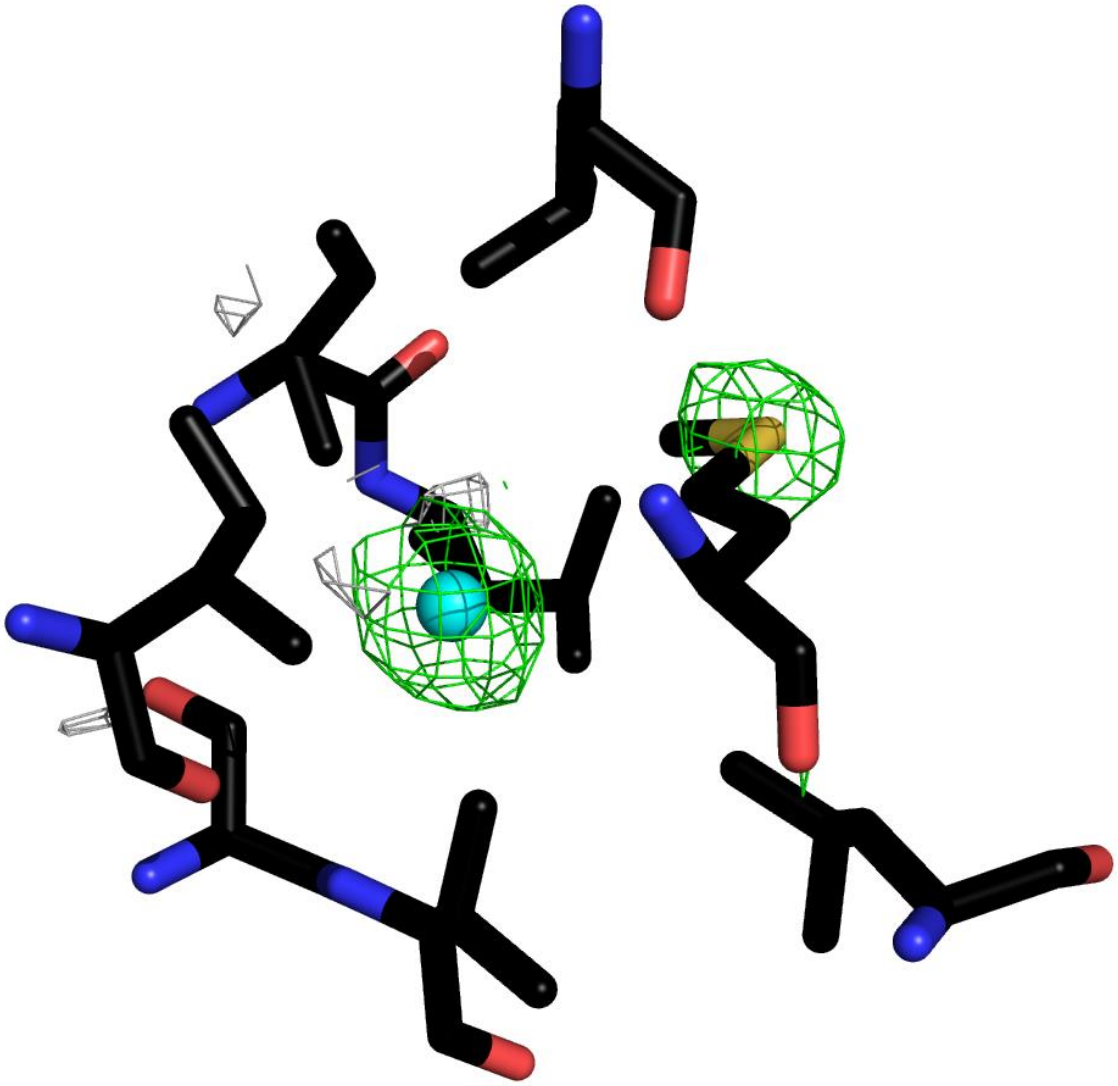


Figure 5.6 HEWL Xe1 Fo-Fc and anomalous maps

A model of likely xenon binding site Xe1 (cyan) and coordinating residues. The Fo-Fc and anomalous maps generated using PHENIX[56] are shown to sigma level 3.0 as grey and green mesh, respectively. The low 0.22 occupancy significantly diminishes the anomalous signal coming from xenon. Also shown is the anomalous signal from neighbouring methionine sulfur (PyMol[62]).

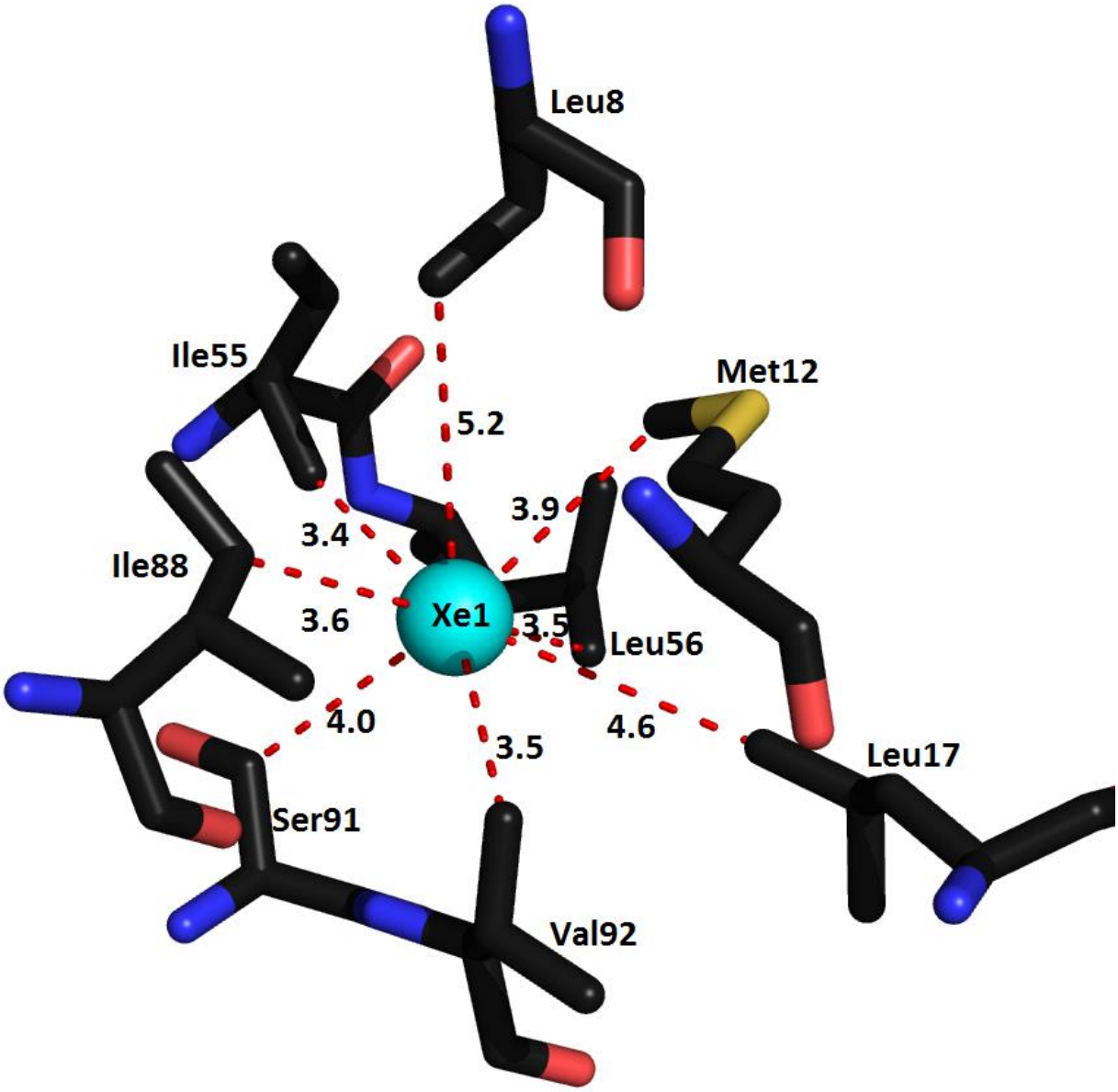


Figure 5.7 HEWL site Xe1 with distances to coordinating residues

A model of likely xenon site Xe1 is shown in cyan with distances to neighbouring amino acids. The site is enclosed in a cavity between 3 alpha helices and is composed of mostly hydrophobic amino acids. All of the nearest conacts within 5Å are to carbon atoms; these kinds of hydrophobic sites are well known to bind xenon[18]. Distances measured to closest contacts indicate bond lengths that are too long for metal binding and there is no metal bond geometry present (PyMol[62]).

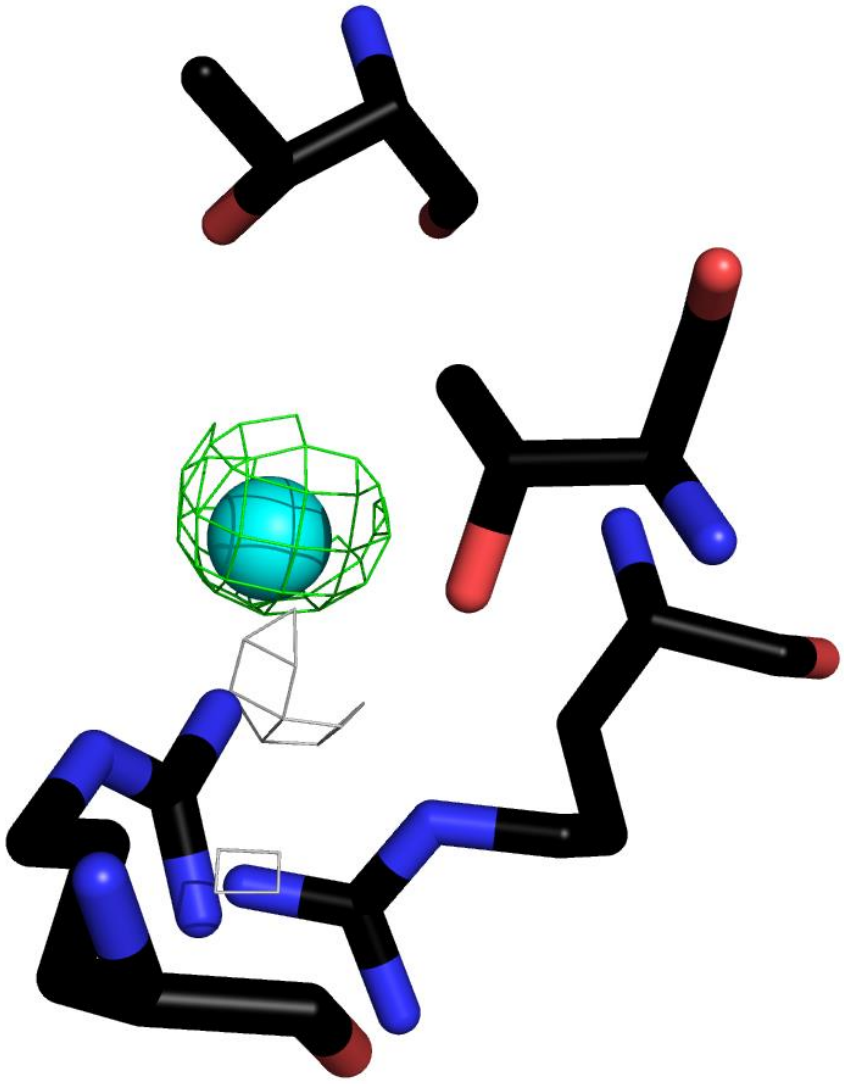


Figure 5.8 HEWL Xe2 Fo-Fc and anomalous maps

A model of likely xenon binding site Xe2 (cyan) and coordinating residues. The Fo-Fc and anomalous maps generated using PHENIX[56] are shown to sigma level 3.0 as grey and green mesh, respectively. Once again, a low 0.14 occupancy significantly diminishes the anomalous signal coming from xenon (PyMol[62]).

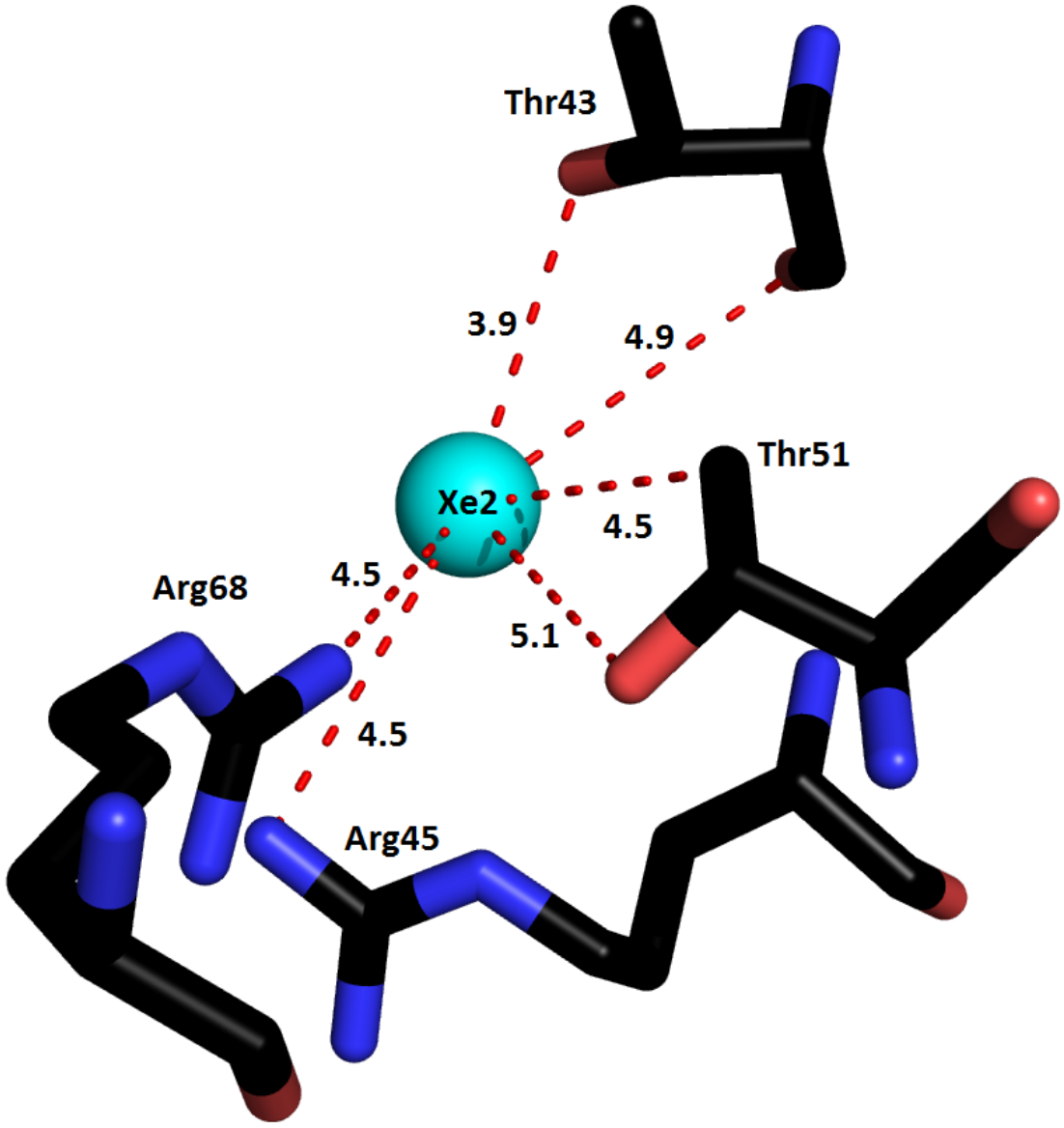


Figure 5.9 HEWL site Xe2 with distances to coordinating residues

A model of likely xenon site Xe2 (cyan) with an occupancy of 0.14 is shown with distances to neighbouring amino acids. This site has a much different character than that of Xe1. Xe2 is located on the edge of the protein trapped between beta sheets of neighbouring protein molecules (Figure 5.10). This site is not a hydrophobic pocket and most of the nearest contacts are to nitrogen and oxygen atoms. The closest contact this time is 3.9Å which is too long for a metal bond and there is no metal bond geometry present (PyMol[62]).

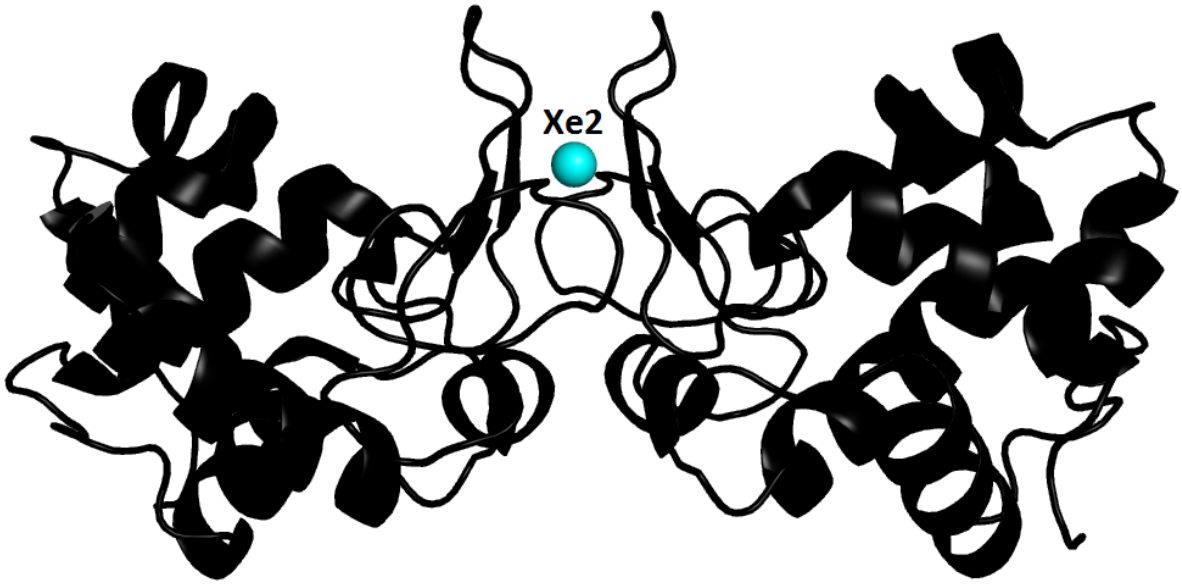


Figure 5.10 HEWL intermolecular xenon Site Xe2

Xenon site Xe2 was the second highest peak in the anomalous maps with an occupancy of 0.14. Unlike Xe1 which is bound between three helices within the protein structure, Xe2 is bound between two beta sheets of adjacent, symmetry-related protein molecules (PyMol[62]).

Thermolysin xenon site Xe1 (cyan) shows a very strong anomalous signal in the anomalous map emanating from the almost full 0.91 occupancy xenon site (Figure 5.11). Like HEWL site Xe1, this site lays buried within the protein, away from the solvent, and in between two alpha helices. The amino acids which make up this site are mostly hydrophobic and include a couple of polar uncharged amino acids as well (Figure 5.12). Once again, there are no contacts within 3.7Å as well as a clear lack of coordination geometry. These aspects further support that xenon binding has taken place.

Thermolysin site Xe6 was selected for further examination due to having both a low refined occupancy of 0.16 (Figure 5.4) as well a relatively high refined f'' value of 3.11 (Figure 5.5). Three other sites Xe2, Xe3, and Xe4 all show coordination geometry indicative of calcium and have refined f'' values in between 3.8-4.5 (Figure 4.2). Xe5 also shows coordination geometry this time indicative of zinc. The weak anomalous signal shown by Xe6 in the

anomalous map (Figure 5.13) is due to the low occupancy and accounts for such a low refined f'' value when compared with xenon site Xe1 which had a refined f'' of 8.24.

The Xe6 site is close by the main site Xe1 only 7.7Å away and closer to the edge of the protein surface. The site is once again mostly hydrophobic and polar with only a single arginine with its side chain facing away from the site. This site does not show coordination geometry and does not appear to be a metal however it differs from other sites observed so far due to the short distance to two neighbouring alanine carboxyl oxygens (Figure 5.14). These oxygens are located 3.0Å and 3.1Å away and although very short, are once again too long for covalent binding[42]. Although a bit small, this sort of smaller pocket may serve as a temporary pathway or binding site for xenon when pressurized. Xenon site Xe6 has a lower affinity for xenon and becomes populated after the main site Xe1 is full. Occurrences of additional xenon sites appearing as a result of increased xenon pressure have been previously observed[18].

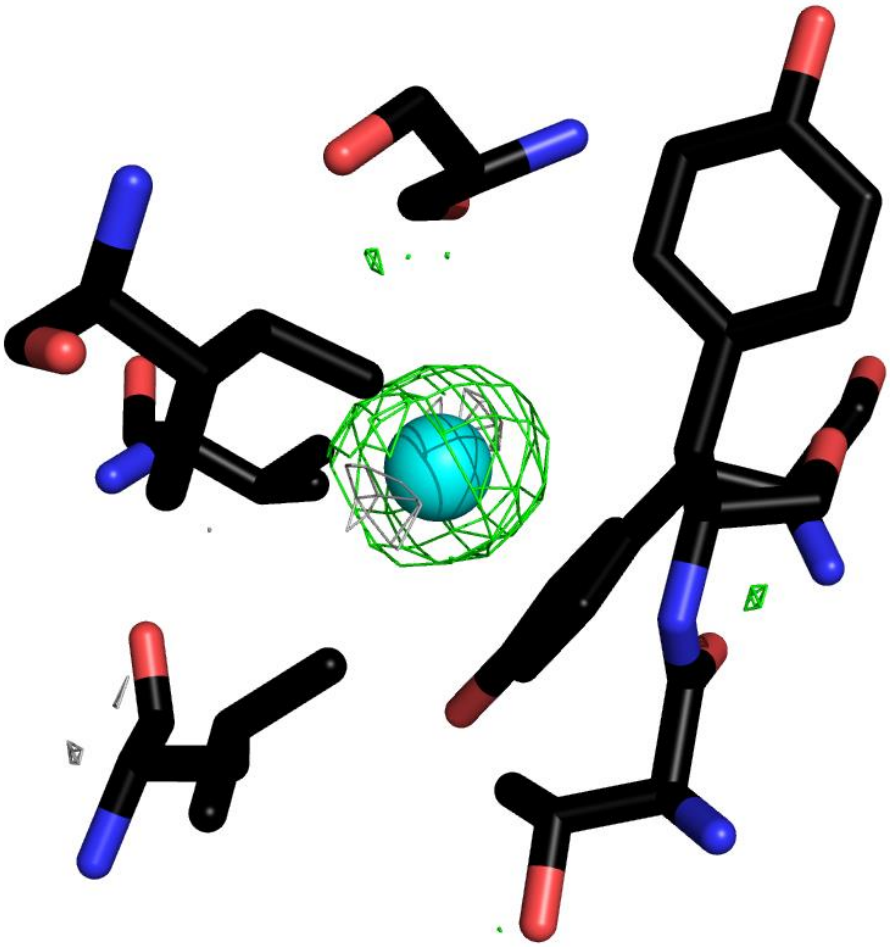


Figure 5.11 Thermolysin Xe1 Fo-Fc and anomalous maps

A model of likely xenon binding site Xe1 (cyan) and coordinating residues. The Fo-Fc and anomalous maps generated using PHENIX[56] are shown to sigma level 3.0 as grey and green mesh, respectively. The almost full 0.91 occupancy site produces a strong anomalous signal and is strongly identified in the anomalous map (PyMol[62]).

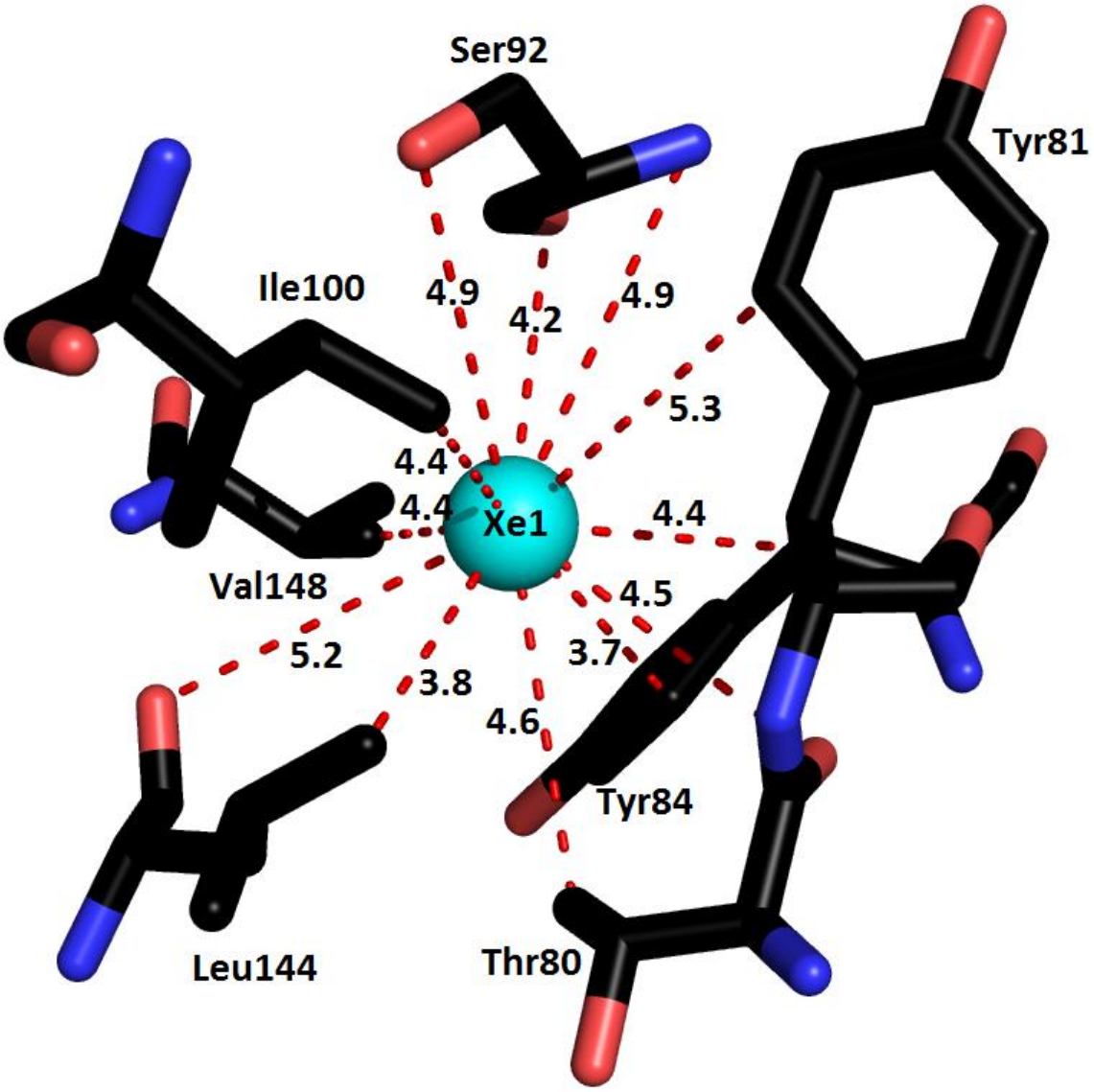


Figure 5.12 Thermolysin site Xe1 with distances to coordinating residues

A model of xenon site Xe1 (cyan) with an occupancy of 0.91 is shown with distances to neighbouring amino acids. The site is enclosed in a cavity between 2 alpha helices and is composed of mostly hydrophobic and neutral amino acids. Most of the nearest contacts within 5Å are to carbon atoms; the shortest contact distance is 3.7Å . The absence of typical metal bond geometry and bond lengths to closest contacts indicate that this is not a metal binding site (PyMol[62]).

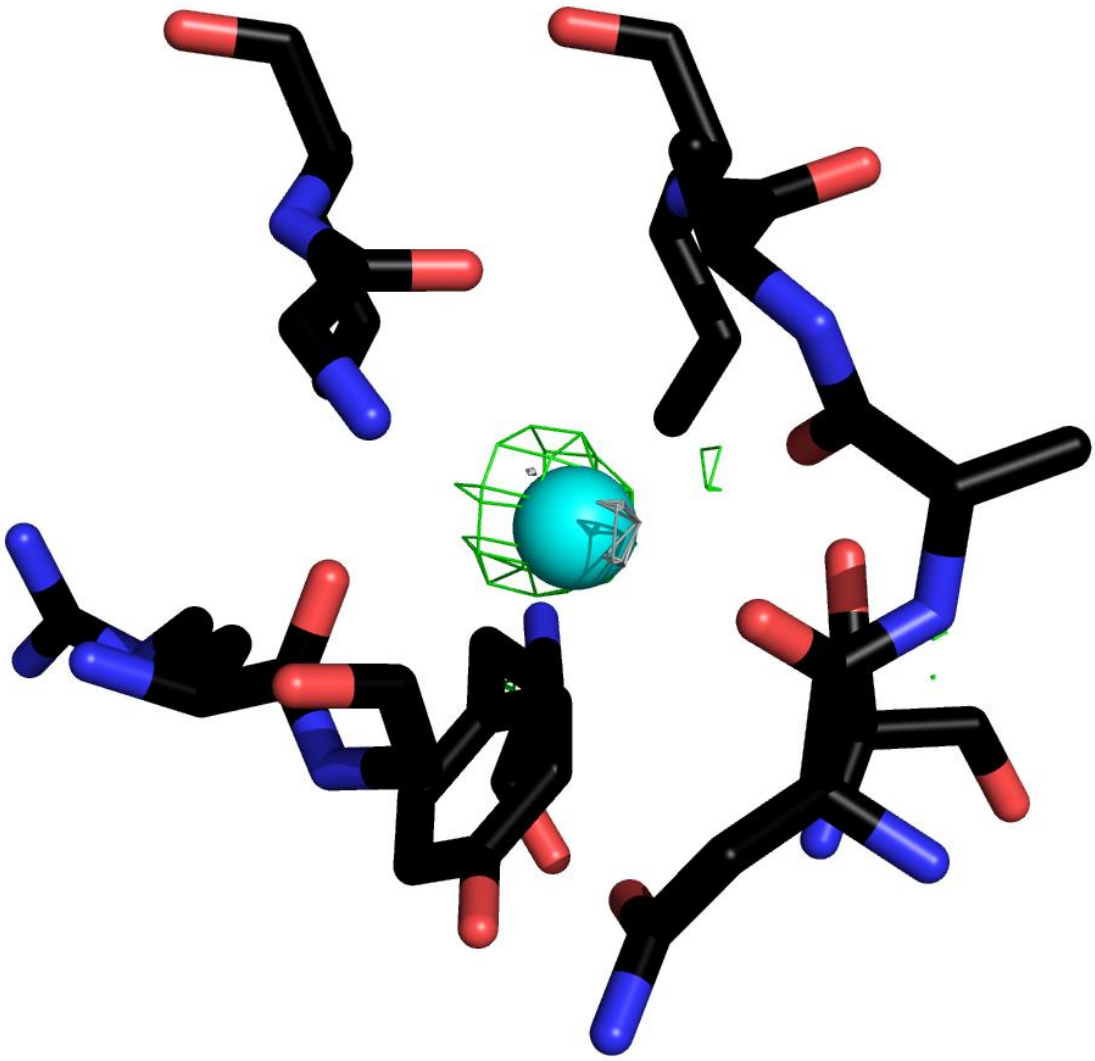


Figure 5.13 Thermolysin Xe6 Fo-Fc and anomalous maps

A model of possible xenon binding site Xe6 (cyan) and coordinating residues. The Fo-Fc and anomalous maps generated using PHENIX[56] are shown to sigma level 3.0 as grey and green mesh, respectively. The low 0.16 occupancy produces a very weak anomalous signal similar to those seen in HEWL (PyMol[62]).

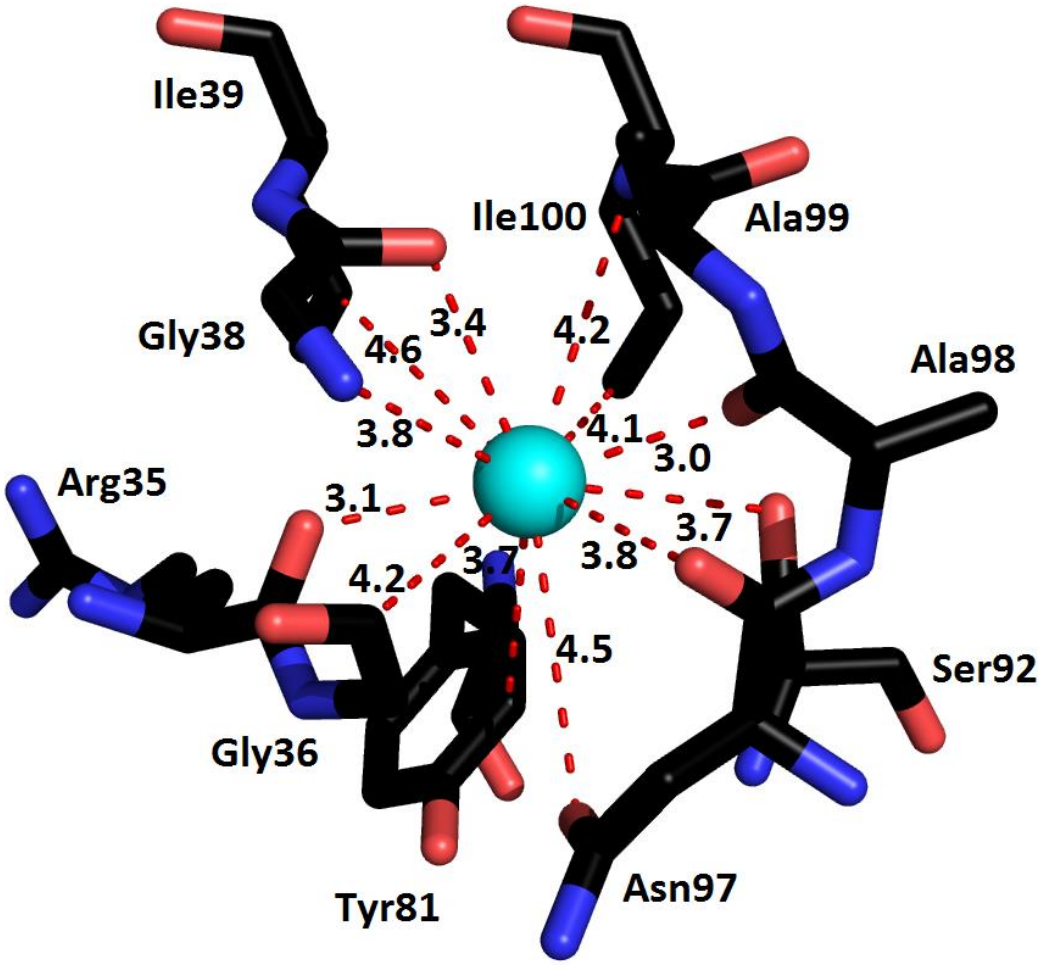


Figure 5.14 Thermolysin site Xe6 with distances to coordinating residues

A model of xenon site Xe6 (cyan) with an occupancy of 0.16 is shown with distances to neighbouring amino acids. Xe6 is located near Xe1 in a coiled region of the protein with 3.0 Å as the shortest contact distance. The absence of typical metal bond geometry and bond lengths to closest contacts indicate that this is not a metal binding site. The absence of chlorides in crystallization and the low number of total peaks in the anomalous maps are strong indicators that this is probably a low occupancy xenon site (PyMol[62]).

The three sites in glucose isomerase all showed very high refined f'' and good potential for being bound xenon (Figure 5.5). Of the three sites, the main site Xe1 with occupancy 0.57 is the only one of the remaining three sites which does not show coordination geometry typical of metal binding. This partially occupied site is strongly identified in the anomalous maps and is located in between two asymmetric units (Figure 5.15). Figures 5.17 & 5.18 show how four glucose isomerase molecules come together during crystallization to produce a binding site for xenon along the solvent channels. Once again the binding site is composed of mainly hydrophobic and polar amino acids as well as two positively charged arginines whose residues are facing away from the site. There are no close contacts within 3.3Å which further supports that this is not metal binding.

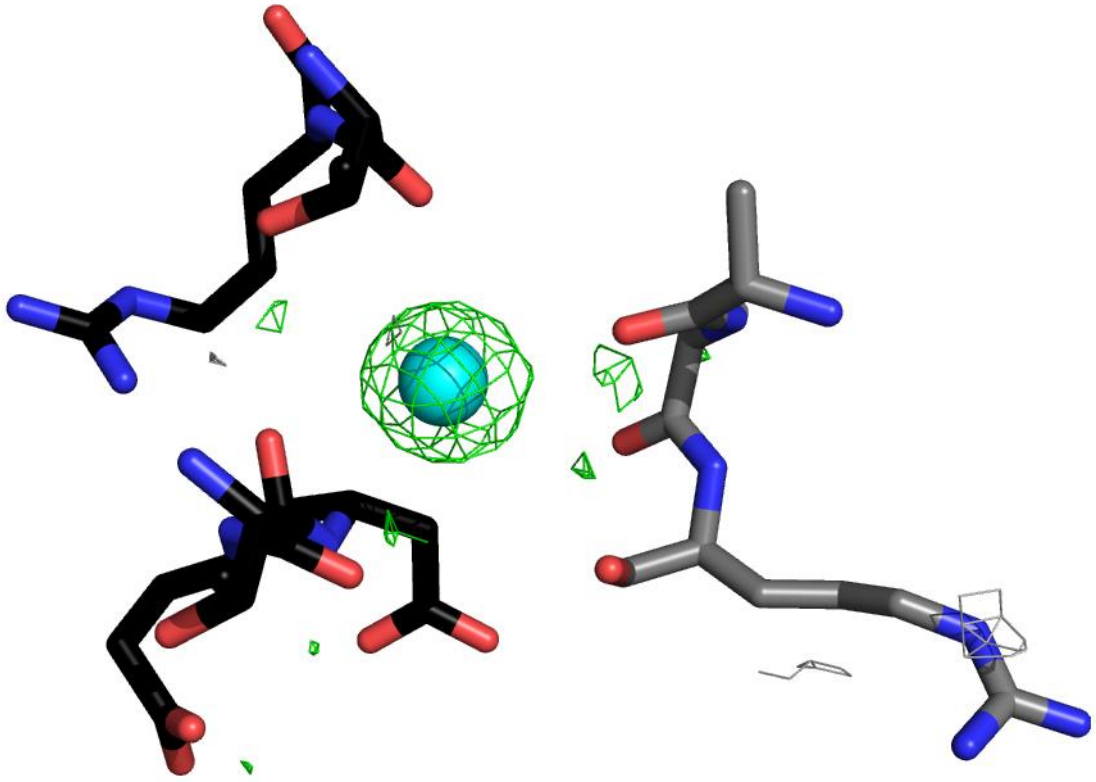


Figure 5.15 Glucose isomerase Xe1 Fo-Fc and anomalous maps

A model of xenon binding site Xe1 (cyan) and coordinating residues. The Fo-Fc and anomalous maps generated using PHENIX[56] are shown to sigma level 3.0 as grey and green mesh, respectively. A mid occupancy of 0.57 produces a strong anomalous signal and is strongly identified in the anomalous map. The portion of the model shown in grey is part of a nearby symmetry related protein which completes the binding site (PyMol[62]).

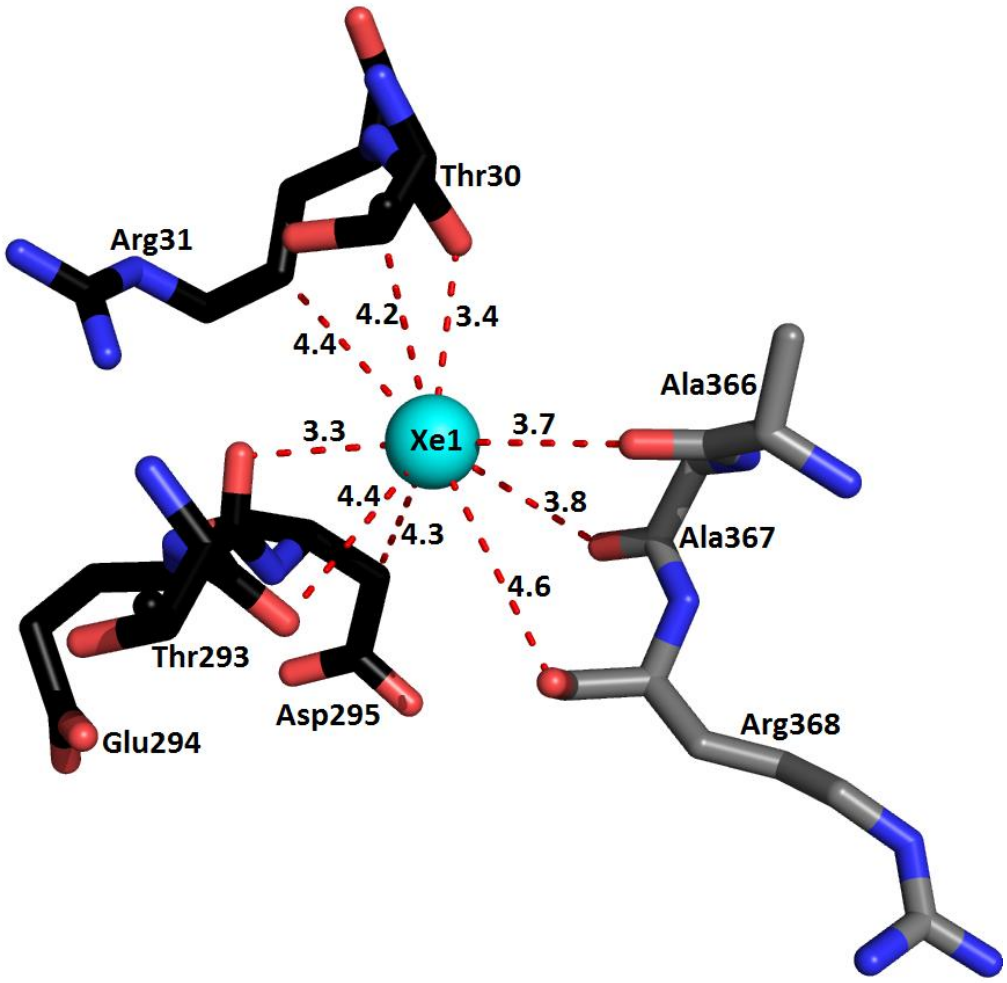


Figure 5.16 Glucose isomerase site Xe1 with distances to coordinating residues

A model of xenon site Xe1 (cyan) with an occupancy of 0.57 is shown with distances to neighbouring amino acids. This site is not a typical hydrophobic pocket and is instead the result of crystal packing which produces a cavity in between the ends of alpha helices of neighbouring protein molecules (Figure 4.14, Figure 4.15). Unexpectedly, there are no really hydrophobic amino acids in this binding site and almost all of the nearest contacts are to oxygens. Visual inspection of this site shows that most of these nearby oxygens belong to the main chain with many side chains facing away from the bound xenon. The absence of typical metal bond geometry and with distances to oxygens all greater than 3.3Å indiate that this is not a metal binding site (PyMol[62]).

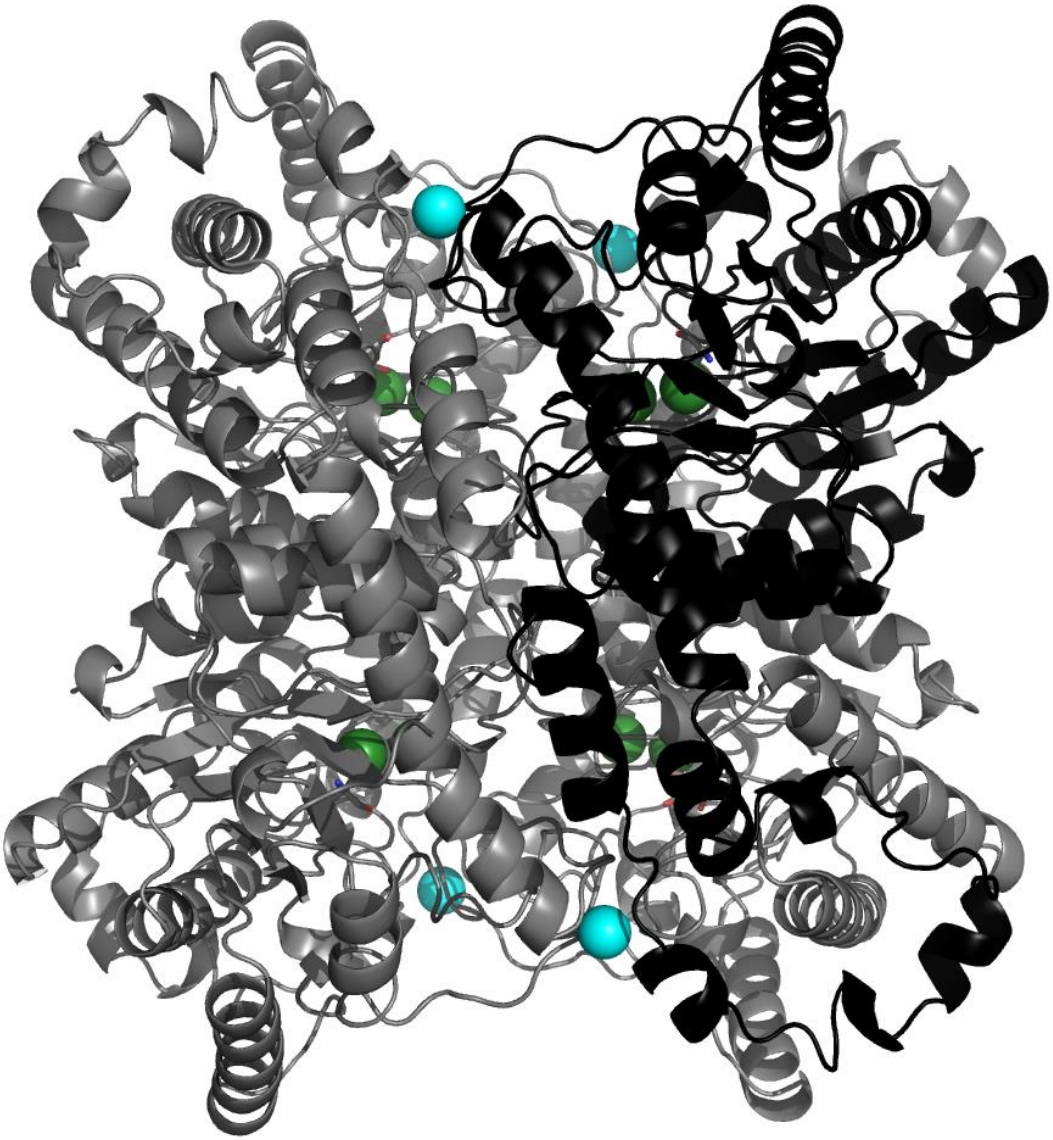


Figure 5.17 Glucose isomerase packing symmetry (front view)

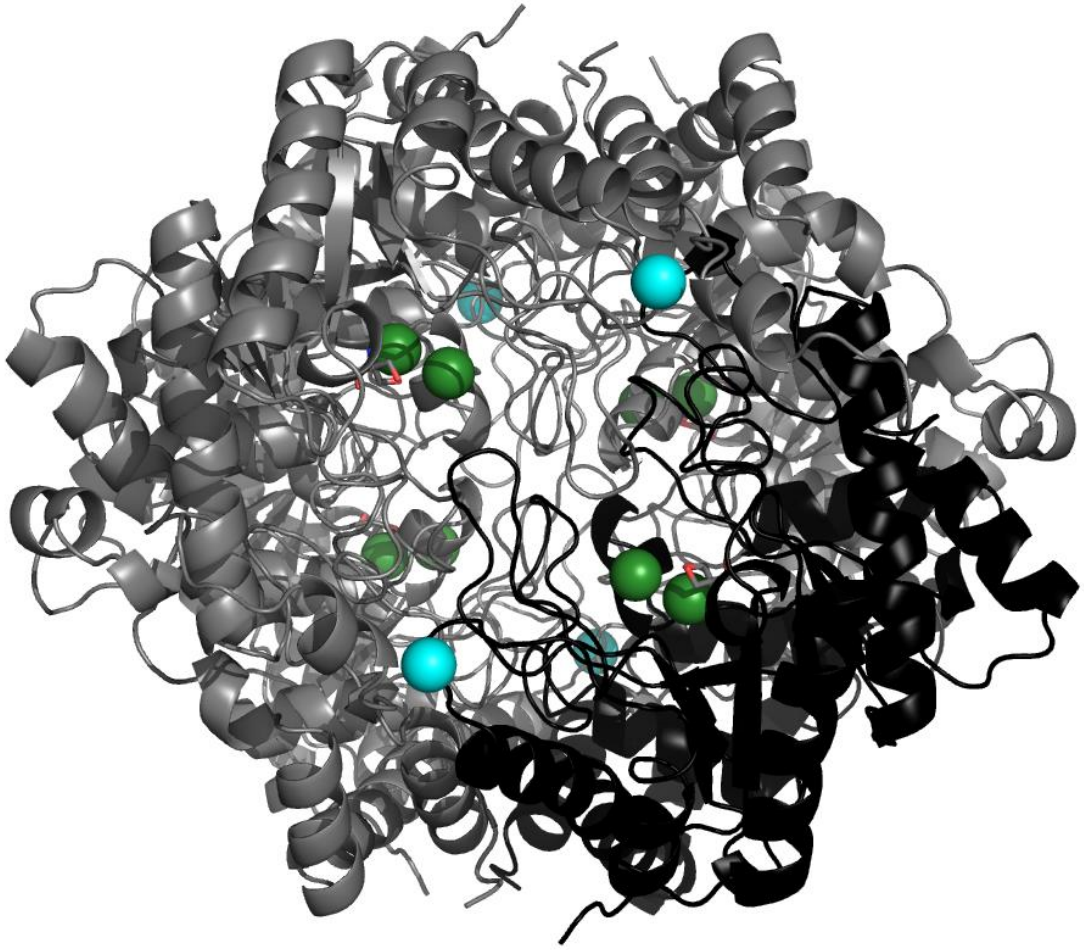


Figure 5.18 Glucose isomerase packing symmetry (top view)

A cartoon ribbon representation of glucose isomerase is shown packed with 3 identical copies shown in grey to form spherical, repeating structures; xenon is once again shown in cyan with magnesium and calcium both in dark green. The front view (Figure 5.17) shows how the xenon (cyan) is bound in between the end of a long alpha helix from one protein and a very short alpha helix of an adjacent protein. Taking a look from the top (Figure 5.1) shows how the ASU packs to form large solvent channels enclosed by the proteins. Within these solvent channels exist small, easily accessible pockets for xenon to travel and bind to (PyMol[62]).

Based on preliminary data (Figures 5.3-5.5), the occupancies and refined f'' values for the thaumatin II xenon sites indicated that further analysis would not be necessary. None of the sites were obviously xenon meaning that xenon binding likely did not occur in this case. The large number of sites located earlier on (Figure 5.3) reflects the fact that there were no xenon sites, metals, or any other strong anomalous scatterers. The strong anomalous signal shown in intial processing (Figure 5.2) is a result of 17 native sulfurs in a 207 amino acid protein. The anomalous signal from these sulfurs only barely stands out above noise in the maps and is the reason for the high number of initial sites (Figure 5.2). Since chloride was not used in crystallization, these remaining non-sulfur sites are most likely simply noise in the anomalous map.

5.8 PDB comparison

The number and variety of PDB models available for each protein varied and sometimes included a xenon derivative. Each xenon site was compared with sites from as many other derivatives as were available. These comparisons are summarized in Tables 5.3-5.5. Comparison for thaumatin II (PDB code: 3AOK[67]) was not included as there was no evidence to suggest xenon binding. Furthermore, comparison showed no presence of any anomalous scatterers in thaumatin II aside from the 17 sulfurs.

Comparison with other structures yielded some interesting results for test protein HEWL (Table 5.3). Most importantly, xenon sites could be confirmed for both Xe1 and Xe2. Interestingly, the higher occupancy site Xe1 and the secondary site Xe2 were reversed with the 1C10[30] xenon derivative showing a higher occupancy in the Xe2 position than in the Xe1 position. Additionally, the 1C10[30] structure showed a third intermediate xenon site which was never observed in any of the HEWL crystals examined. Study of other available structures[8,30,68] also showed that the large number of initial peaks observed was largely due to chloride ions which bind readily along solvent channels. A large number of chloride ions were observed in each of the structures in which chloride was used as an ingredient in crystallization. It is also interesting to note that the hydrophobic main site Xe1 can also

bind noble gas krypton while the secondary site Xe2 has displaced solvent which would otherwise have been there.

Comparison of thermolysin structures (Table 5.4) supported xenon binding for site Xe1. Metal binding sites Xe2-Xe4 were reported as calcium ions with Xe5 reported as a zinc ion[69,70]. Interestingly, in the other two thermolysin structures examined, both reported four calcium ions while only three calcium ions were ever observed with nothing appearing in the anomalous map. Xe6 is interesting and unusual as discussed previously. Comparison with the two other structures, including a xenon derivative, shows that this site is usually occupied by solvent. The lack of chloride during crystallization strongly suggests that this is a low occupancy xenon site; observation of this low occupancy xenon site has not yet been reported. Final inspection of the anomalous maps shows that site Xe8 is not real which is expected given the low occupancy and refined f''. It is uncertain as to why this peak was not removed automatically during previous model building and refinement. It may have to do with this site's location in a particularly disordered region near a zinc ion in the active site.

Comparison of glucose isomerase structures[71] (Table 5.5) was useful in identifying the two metals Xe2 and Xe3 as calcium and magnesium, respectively. The main site Xe1 has replaced water in the native structure and has replaced a chloride ion in the available krypton structure. Observation of xenon site Xe1 has not yet been reported. It is interesting to note that the two krypton sites are unique from the xenon site unlike in the case of HEWL where xenon and krypton had one site in common. It is also noteworthy that, in the absence of xenon, chloride has been observed in one case to fill the cavity along the solvent channel. Chloride ions are known to prefer some of the same sites as xenon[49] which suggests that xenon may have to compete with chloride ions, when present, resulting in sites which may be partially occupied by both. In the case of HEWL Xe2, the low xenon occupancy and the below theoretical refined f'' value could be the result of software struggling to make sense of a site which is partially occupied by two very different elements.

Table 5.3 HEWL PDB comparison

Data * (Xe)	1C10[30] * (Xe)	1LZ8[8] * (Cl ,S)	1QTK[30] * (Kr)	1VAT[68] (I)	*Cl
Xe 1 (0.22)	Xe 152 (0.15)	-	Kr 150 (0.49)	-	
Xe 2 (0.14)	Xe 150 (0.32)	H2O	H2O	H2O	
Xe 3 (0.33)	-	-	-	-	
Xe 6 (0.22)	-	H2O	-	H2O	
Xe 7 (0.35)	Cl 160	Cl 201	Cl 160	I 301 (0.22)	
Xe 13 (0.30)	Cl 165	Cl 205	H2O	H2O	
Xe 15 (0.17)	Cl 163	Cl 204	Cl 163	H2O	
Xe 16 (0.22)	Cl 161	Cl 202	Cl 161	H2O	
Xe 17 (0.35)	Cl 162	Cl 203	Cl 162	H2O	
Xe 19 (0.28)	-	Cl 206	-	H2O	
Xe 21 (0.27)	H2O	Cl 207	H2O	H2O	
Xe 22 (0.24)	-	-	-	-	
Xe 23 (0.21)	H2O	H2O	H2O	H2O	
H2O	Xe 151 (0.28)	H2O	H2O	H2O	
-	Cl 164	H2O	H2O	H2O	
H2O	-	H2O	-	I 300 (0.36)	

Several PDB structures are compared with the data. Along with their PDB codes, in parentheses is the type of derivative for each structure. An asterisk next to the PDB code indicates that chloride was used as an ingredient in crystallization. Beside each element is the chain number for the element along with occupancy in parentheses when applicable.

Table 5.4 Thermolysin PDB comparison

My Data (Xe)	3LS7[69] (Xe)	2TLX[70] (native)
Xe 1 (0.91)	Xe 501 (0.81)	-
Xe 2 (0.35)	Ca 604	Ca 323
Xe 3 (0.41)	Ca 601	Ca 325
Xe 4 (0.43)	Ca 602	Ca 326
Xe 5 (0.66)	Zn 701	Zn 322
Xe 6 (0.16)	H2O	H2O
H2O	Ca 603	Ca 624

Two PDB structures are compared with the data. Along with their PDB codes, in parentheses is the type of derivative for each structure. Beside each element is the chain number for the element along with occupancy in parentheses when applicable.

Table 5.5 Glucose isomerase PDB comparison

My Data ** (Xe)	1O1H[71] * ** (Kr)	1MNZ[71] * ** (Native)	*Cl ** Mg
Xe 1 (0.57)	Cl 1010	H2O	
Xe 2 (0.48)	Ca 1001	Ca 390	
Xe 3 (0.47)	Mg 1003	Mg 389	
-	Kr 1006 (0.20)	-	
-	Kr 1008 (0.40)	-	

Two PDB structures are compared with the data. Along with their PDB codes, in parentheses is the type of derivative for each structure. An asterisk next to the PDB code indicates that chloride was used as an ingredient in crystallization while a double asterisk indicates the use of magnesium. Beside each element is the chain number for the element along with occupancy in parentheses when applicable.

5.9 Protein Summary

Xenon was observed to bind to three out of four proteins as summarized in Table 5.6. In all cases, there was sufficient anomalous signal to locate the anomalous scatterers even for thaumatin II which did not bind xenon. Anomalous signal came not only from xenon but from a number of other anomalous scatterers and caused difficulty in accurately assessing the phasing power of xenon alone. Thaumatin II and HEWL were not ideal cases. The absence of any major anomalous scatterers and a high number of sulfurs, and chlorides in the case of HEWL, produced maps with upwards of 20 small peaks. The large number of peaks which originate from these light anomalous scatterers made it difficult to identify low occupancy xenon binding sites. Due to the large number of anomalous scatterers in these cases, it is difficult to know how significantly xenon contributed to the anomalous signal.

Thermolysin was the cleanest example. With only two sulfurs and a few metals, the high occupancy xenon site played a much more important role in the phasing of this protein. The location of the second xenon binding site would also not have been possible as it would be obscured by a large number of small peaks. Glucose isomerase was another very good

example with only a couple of metals and 9 sulfurs. The approximately half-filled xenon site and two metals provided the majority of the anomalous signal for the largest of the four proteins. In each case where xenon binding was observed, xenon was not the sole contributor to anomalous signal.

Based on the studies of thermolysin and glucose isomerase, it is probable that a single half occupied xenon site is sufficient to phase proteins of this size although a case of xenon being the sole contributor of anomalous signal was not observed in this study. Ribbon cartoon representations of each proteins studied along with anomalous scatterers are depicted in Figures 5.19-5.22.

Table 5.6 Summary of results

	Lysozyme	Thermolysin	Glucose Isomerase	Thaumatin II
Weight (kDa)	14.3	34.3	43.2	22.3
Time (min)	5	5	5	5
Pressure (psi)	300	300	500	300
Multiplicity	10.1	13.2	5	10.3
FOM	0.44	0.475	0.371	0.429
# xenon sites	2	2	1	0
Occupancy	0.22	0.91	0.57	
	0.13	0.16		
# other anomalous scatterers	20	6	11	17

A summary of xenon binding. Non-xenon anomalous scatterers are a combination of metals, chlorides, and sulfurs. The lower FOM value for glucose isomerase is likely due to the lower multiplicity more than anything else.

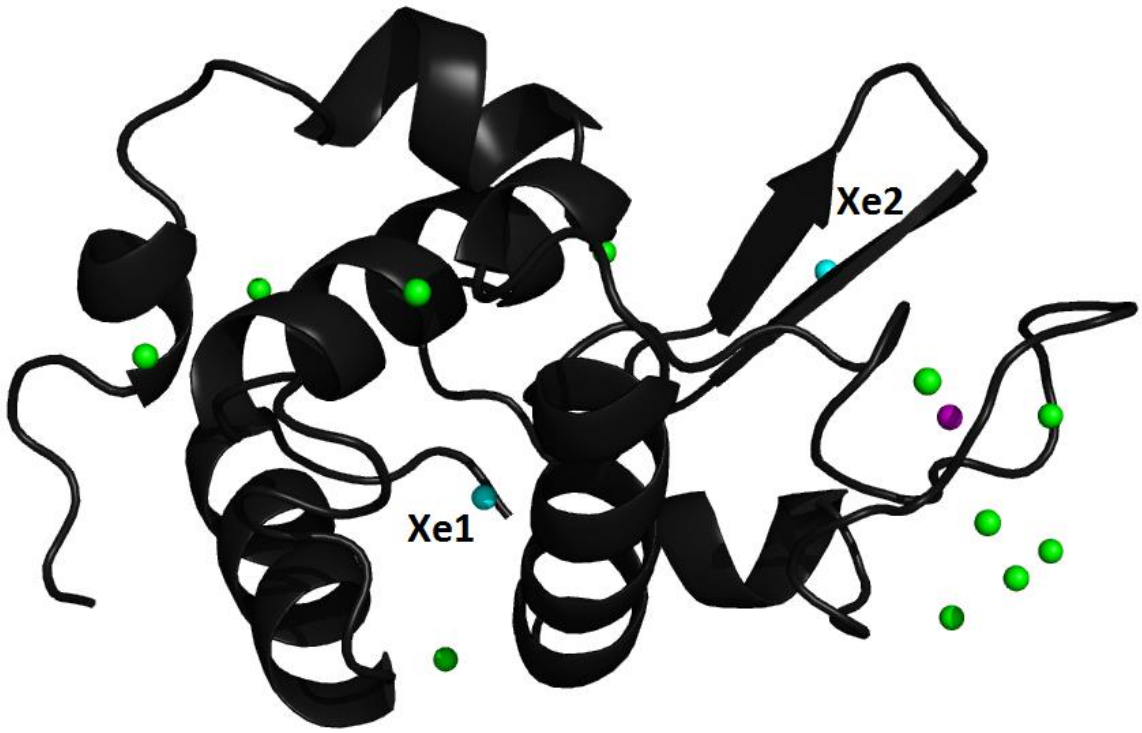


Figure 5.19 Ribbon cartoon representation of HEWL with bound xenon

Bound xenon is shown in cyan along with a single sodium ion shown in purple and chloride ions shown in light green. Diffraction data were collected on the 08B1-1 (CMCF-BM) beamline from a single HEWL crystal. Anomalous signal from two low-occupancy xenon sites, 10 native sulfurs, and 11 chloride ions at 7keV provided sufficient anomalous signal to solve the phase problem (PyMol[62]).

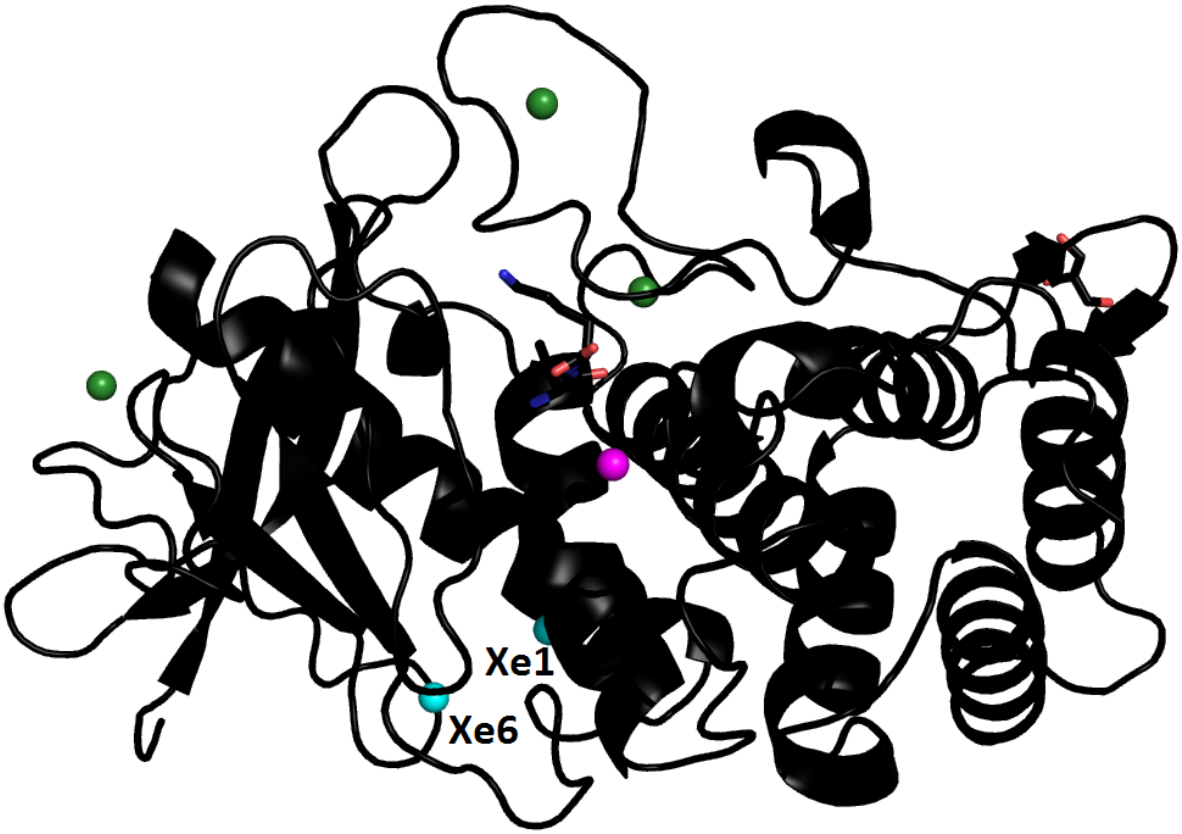


Figure 5.20 Ribbon cartoon representation of thermolysin with bound xenon

Bound xenon is shown in cyan along with three calcium ions shown in dark green and zinc shown in purple. Also shown in the active site is a stick representation of the amino acids valine and lysine. Diffraction data were collected on the 08B1-1 (CMCF-BM) beamline from a single thermolysin crystal. Anomalous signal from one high-occupancy and one low-occupancy xenon site, 2 native sulfurs, 3 calcium ions, and a zinc at 7keV provided sufficient anomalous signal to solve the phase problem (PyMol[62]).

Figure 5.21 Ribbon cartoon representation of glucose isomerase with bound xenon

Bound xenon is shown in cyan with one calcium ion and one magnesium ion both shown in dark green. Also shown is a stick representation of tris which was used in the crystallization conditions. Diffraction data were collected on the 08B1-1 (CMCF-BM) beamline from a single glucose isomerase crystal. Anomalous signal from a single xenon site, 9 native sulfurs, one calcium ion, and one magnesium ion at 7keV provided sufficient anomalous signal to solve the phase problem (PyMol[62]).

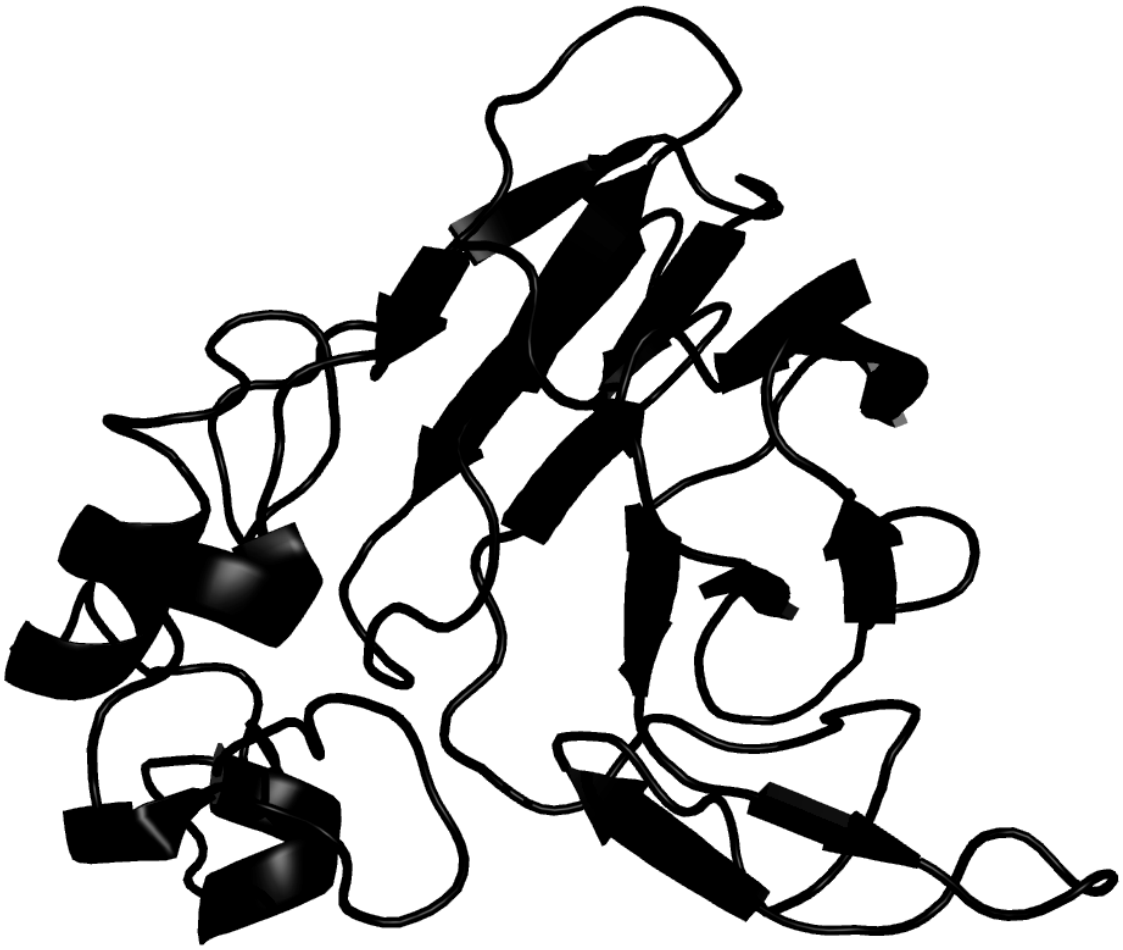


Figure 5.22 Ribbon cartoon representation of thaumatin II structure

Xenon was not found to bind to thaumatin II. Thaumatin II contains no native or bound metals. Diffraction data were collected on the 08B1-1 (CMCF-BM) beamline from a single thaumatin II crystal. The anomalous signal from 17 native sulfurs at 7keV provided sufficient anomalous signal to solve the phase problem (PyMol[62]).

6.0 Conclusions

Xenon derivative crystals were produced using xenon gas pressures in between 300-500 psi for 5 minutes. Producing derivative crystals in this manner, along with data collection and processing strategies outlined in this document, will serve as a guide for students and researchers who wish to attempt xenon derivatization on their own crystals. The study of HEWL was one of the more difficult proteins investigated and resulted in a strategy which can easily be implemented on other proteins. Results varied but included low-occupancy xenon binding and, in one case, no xenon binding.

Xenon derivatization is a physical method that can be used to produce derivatives when production of a SeMet derivative is not possible. A single partially occupied xenon site, as long as it was not too low, should be sufficient to phase proteins of the size investigated in this study. As will often be the case, xenon may provide the majority of the anomalous signal but it will be simply a contribution to a signal which may have been sufficient absent the xenon.

6.1 Pros & Cons

The use of xenon for producing derivatives is convenient due to the ease with which derivatives may be produced. In addition to requiring a short 5 minutes to prepare, the nature of the mechanical binding means that there is minimal effect on both the protein and crystal structures. The non-covalent, weak-energy van der Waals interactions which exist between xenon and proteins makes the process of xenon binding completely reversible; this may be demonstrated by releasing xenon gas pressure and allowing xenon to diffuse out of the crystal leaving the crystal unchanged. The minimal change caused by xenon binding leads to a high degree of isomorphism with native crystals. Even though isomorphous differences are not used in SAD experiments, a structure which is as close to the native structure as possible should be obtained whenever possible.

The phasing power and nature of xenon is advantageous when compared to other heavy atoms used in SAD. Xenon is much heavier than other SAD anomalous scatterers (Figure 2.8). Its large size produces a very strong anomalous signal and does so without chemical

alteration characteristic of salt soaks and SeMet. In addition to global and local changes which occur due to radiation damage, heavy atoms are damaged by radiation. Larger atoms are able to absorb a greater amount of radiation[23] making xenon superior to other SAD scatterers including selenium.

Furthermore, it has been previously demonstrated[18,32] that changing the xenon gas pressure may result in additional xenon sites thus producing derivatives which are complementary to other types of derivatives. Xenon is again complementary to other derivatives since xenon has affinity for apolar and/or hydrophobic environments and likely binds to sites that are different from those of standard heavy atoms.

As with many other derivatives, it is not certain whether or not xenon binding has taken place until after data has been collected and undergone preliminary processing. After a lengthy and high redundancy data collection, xenon binding may not have occurred or perhaps the quality of the data was inadequate to measure the anomalous signal. This leads to lost time as there are no guarantees of xenon binding, occupancy of binding, number of sites, effects of radiation damage, crystal quality, and diffraction data quality.

6.2 Future Work

This document will serve as a guide for all those who would attempt xenon derivatization on their crystals. At the time of this writing, an attempt has already been made to perform xenon SAD on a protein crystal for Dr. Marie Fraser of the Department of Biological Sciences at the University of Calgary. Unfortunately, the attempt was not successful in generating starting phases as the protein either had not bound xenon or the data quality was too poor.

Possible further work includes the use of a helium box whereby absorption of scattered x-rays by air is reduced by replacing air with helium[72,73]. This could be used as a last resort in cases where xenon is thought to have bound but where the anomalous signal is very weak due to low occupancy xenon sites. The use of a helium box may also be beneficial in cases of very large proteins having an insufficient number and occupancy of xenon sites for

phasing to be successful. In particular, a helium box could be used in conjunction with a lower x-ray energy thus maximizing the anomalous signal from xenon by approaching an absorption edge without sacrificing anomalous signal lost to air absorption[72,73].

Two xenon sites in this study were observed to be the result of the packing of asymmetric units in the crystal structure. Previous studies have shown that controlled crystal hydration may lead to rearrangement of the domains in a crystal[74]. In one case, the cycling of hydration and dehydration caused one protein, upon dehydration, to pack more efficiently into a higher symmetry space group[75]. A case like this would be ideal for studying the generation of xenon binding sites by controlled crystal hydration.

7.0 References

1. Hendrickson, W. a. Anomalous diffraction in crystallographic phase evaluation. *Q. Rev. Biophys.* **47,** 49–93 (2014).

2. Weinert, T. *et al.* Fast native-SAD phasing for routine macromolecular structure determination. *Nat. Methods* **12,** (2014).

3. Mills, D. M. *Third-Generation Hard X-ray Synchrotron Radiation Sources*. (John Wiley & Sons, Inc., 2002).

4. Grochulski, P. *et al.* Canadian macromolecular crystallography facility: a suite of fully automated beamlines. *J. Struct. Funct. Genomics* **13,** 49–55 (2012).

5. Rupp, B. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. (Garland Science, 2009).

6. Hendrickson, W. a, Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9,** 1665–72 (1990).

7. Hendrickson, W. A. & Teeter, M. M. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature* **290,** 107 (1981).

8. Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. Can anomalous signal of sulfur become a tool for solving protein crystal structures? *J. Mol. Biol.* **289,** 83–92 (1999).

9. Usón, I. & Sheldrick, G. M. Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9,** 643–648 (1999).

10. Karle, J. & Hauptman, H. A theory of phase determination for the four types of non-centrosymmetric space groups 1 P 222, 2 P 22, 3 P 1 2, 3 P 2 2. *Acta Crystallogr.* **9,** 635–651 (1956).

11. Massa, W. *Crystal Structure Determination*. (Springer, 2004).

12. Sayre, D. The squaring method: a new method for phase determination. *Acta Crystallogr.* **5,** 60–65 (1952).

13. Rhodes, G. *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*. (Elsevier Inc., 2006).

14. Boggon, T. J. & Shapiro, L. Screening for phasing atoms in protein crystallography. *Structure* **8,** R143–9 (2000).

15. Hendrickson, W. a. Maturation of MAD phasing for the determination of macromolecular structures. *J. Synchrotron Radiat.* **6,** 845–851 (1999).

16. Abendroth, J. *et al.* SAD phasing using iodide ions in a high-throughput structural genomics environment. *J. Struct. Funct. Genomics* **12,** 83–95 (2011).

17. Dauter, Z., Li, M. & Wlodawer, A. Practical experience with the use of halides for phasing macromolecular structures : a powerful tool for structural genomics. *Acta Crystallogr. Sect. D* **57,** 239–249 (2001).

18. Schiltz, M., Fourme, R. & Prangé, T. Use of noble gases xenon and krypton as heavy atoms in protein structure determination. *Methods Enzymol.* **374,** 83–119 (2003).

19. Quillin, M. L. & Matthews, B. W. Generation of noble-gas binding sites for crystallographic phasing using site-directed mutagenesis. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58,** 97–103 (2001).

20. Dauter, Z. Carrying out an optimal experiment. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 389–92 (2010).

21. Vekhter, Y. Improving experimental phasing: the role of strongest reflections. *Acta Crystallogr. D. Biol. Crystallogr.* **61,** 899–902 (2005).

22. Shimizu, N., Hirata, K., Hasegawa, K., Ueno, G. & Yamamoto, M. Dose dependence of radiation damage for protein crystals studied at various X-ray energies. *J. Synchrotron Radiat.* **14,** 4–10 (2007).

23. Holton, J. M. A beginner's guide to radiation damage. *J. Synchrotron Radiat.* **16,** 133–42 (2009).

24. Pflugrath, J. Macromolecular cryocrystallography - methods for cooling and mounting protein crystals at cryogenic temperatures. *Methods* **34,** 415–423 (2004).

25. Schoenborn, B. P., Herman, W. C. & Kendrew, J. C. Binding of Xenon to Sperm Whale Myoglobin. *Nature* **207,** 28–30 (1965).

26. Schoenborn, B. P. Binding of Xenon to Horse Haemoglobin. *Nature* **208,** 760–762 (1965).

27. Schoenborn, B. P. & Nobbs, C. L. The Binding of Xenon to Sperm Whale Deoxymyoglobin. *Mol. Pharmacol.* **2,** 495–498 (1966).

28. Tilton, R. F., Kuntz, I. D. & Petsko, G. a. Cavities in proteins: structure of a metmyoglobin-xenon complex solved to 1.9 A. *Biochemistry* **23,** 2849–57 (1984).

29. Schiltz, M., Fourme, R., Broutin, I. & Prangé, T. The catalytic site of serine proteinases as a specific binding cavity for xenon. *Structure* **3,** 309–16 (1995).

30. Prangé, T. *et al.* Exploring Hydrophobic Sites in Proteins With Xenon or Krypton. *Proteins Struct. Funct. Genet.* **30,** 61–73 (1998).

31. Cohen, A., Ellis, P., Kresge, N. & Soltis, S. M. MAD phasing with krypton. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57,** 233–238 (2001).

32. Stowell, M. H. B., Soltis, S. M., Kisker, C., Peters, J. W. & Schindelin, H. A Simple Device for Studying Macromolecular Crystals Under Moderate Gas Pressures (0.1-10MPa). *J. Appl. Crystallogr.* **29,** 608–613 (1996).

33. Schiltz, M., Prangé, T. & Fourme, R. On the preparation and X-ray data collection of isomorphous xenon derivatives. *J. Appl. Crystallogr.* **27,** 950–960 (1994).

34. Kim, C. U., Hao, Q. & Gruner, S. M. Solution of protein crystallographic structures by high-pressure cryocooling and noble-gas phasing. *Acta Crystallogr. D. Biol. Crystallogr.* **62,** 687–94 (2006).

35. Schiltz, M. *et al.* High-pressure krypton gas and statistical heavy-atom refinement: a successful combination of tools for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.* **53,** 78–92 (1997).

36. Antropoff, A. von. No Title. *Z.Elektrochem* 269–308 (1919).

37. Wilhelm, E., Battino, R. & Wilcock, R. J. No Title. *Chem. Rev.* 219–259 (1977).

38. Kennan, R. P. & Pollack, G. L. No Title. *J. Chem. Phys.* 6569–6579 (1990).

39. Sauer, O., Schmidt, a. & Kratky, C. Freeze-Trapping Isomorphous Xenon Derivatives of Protein Crystals. *J. Appl. Crystallogr.* **30,** 476–486 (1997).

40. Schiltz, M. *et al.* Protein Crystallography at Ultra-Short Wavelengths: Feasibility Study of Anomalous-Dispersion Experiments at the Xenon K-edge. *J. Synchrotron Radiat.* **4,** 287–97 (1997).

41. Allen, F. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. Sect. B Struct. Sci.* 380–388 (2002). at <http://scripts.iucr.org/cgi-bin/paper?an0610>

42. Moody, J. A Decade of Xenon Chemistry. *J. Chem. Educ.* **51,** 828–30 (1962).

43. Israelachvili, J. *Intermolecular & Surface Forces*. (Academic Press, 1992).

44. Weiss, M. S., Sicker, T., Djinovic-Carugo, K. & Hilgenfeld, R. On the routine use of soft X-rays in macromolecular crystallography. *Acta Crystallogr. Sect. D* **57,** 689–695 (2001).

45. Mueller-Dieckmann, C., Panjikar, S., Tucker, P. a & Weiss, M. S. On the routine use of soft X-rays in macromolecular crystallography. Part III. The optimal data-collection wavelength. *Acta Crystallogr. D. Biol. Crystallogr.* **61,** 1263–72 (2005).

46. Weiss, M. S., Sicker, T., Hilgenfeld, R. & D-, J. Soft X-rays , High Redundancy , and Ways & Means Proper Scaling : A New Procedure for Automated Protein Structure Determination via SAS. *Structure* **9,** 771–777 (2001).

47. Mueller-Dieckmann, C. *et al.* On the routine use of soft X-rays in macromolecular crystallography. Part II. Data-collection wavelength and scaling models. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 28–38 (2004).

48. Abendroth, J. *et al.* SAD phasing using iodide ions in a high-throughput structural genomics environment. *J. Struct. Funct. Genomics* **12,** 83–95 (2011).

49. Dauter, Z., Dauter, M. & Dodson, E. Jolly SAD. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58,** 494–506 (2002).

50. Vitali, J., Robbins, a. H., Almo, S. C. & Tilton, R. F. Using xenon as a heavy atom for determining phases in sperm whale metmyoglobin. *J. Appl. Crystallogr.* **24,** 931–935 (1991).

51. Stackelberg, M. vo. & Muller, H. R. No Title. *Z. Elektrochem* **58,** 25 (1954).

52. Read, R. J. & McCoy, A. J. Using SAD data in Phaser. *Acta Crystallogr. D. Biol. Crystallogr.* **67,** 338–44 (2011).

53. Kim, C. U., Barstow, B., Tate, M. W. & Gruner, S. M. Evidence for liquid water during the high-density to low-density amorphous ice transition. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 4596–600 (2009).

54. Fodje, M. *et al.* MxDC and MxLIVE: software for data acquisition, information management and remote access to macromolecular crystallography beamlines. *J. Synchrotron Radiat.* **19,** 274–80 (2012).

55. Kabsch, W. Automatic indexing of rotation diffraction patterns. *J. Appl. Crystallogr.* **21,** 67–72 (1988).

56. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 213–21 (2010).

57. Grosse-Kunstleve, R. W. & Adams, P. D. Substructure search procedures for macromolecular structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **59,** 1966–1973 (2003).

58. Terwilliger, T. C. *et al.* Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D. Biol. Crystallogr.* **65,** 582–601 (2009).

59. Blow, D. M. & Crick, F. H. C. The treatment of errors in the isomorphous replacement method. *Acta Crystallogr.* **12,** 794–802 (1959).

60. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D. Biol. Crystallogr.* **64,** 61–9 (2008).

61. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 486–501 (2010).

62. DeLano, W. L. The PyMol Molecular Graphics System. (2002). at <http://www.pymol.org>

63. Private communication reference.

64. Metal coordination of several divalent cations. at <www.iucr.org/__data/iucr/cong/17/iucr/abstracts/abstracts/S0250.html>

65. Transition Metals and Coordination Chemistry. at <http://depts.washington.edu/chemcrs/bulkdisk/chem162A_win10/notes_Lecture_Mar 1 2010.pdf>

66. Metal-Donor atom target distances. at <http://tanna.bch.ed.ac.uk/newtargs_06.html>

67. Masuda, T., Ohta, K., Tani, F., Mikami, B. & Kitabatake, N. Crystal structure of the sweet-tasting protein thaumatin II at 1.27Å. *Biochem. Biophys. Res. Commun.* **410,** 457–460 (2011).

68. Takeda, K. *et al.* Multi-wavelength anomalous diffraction method for I and Xe atoms using ultra-high-energy X-rays from SPring-8. *J. Appl. Crystallogr.* **37,** 925–933 (2004).

69. Behnen, J., Krueger, H., Toker, L., Brumshtein, B., Silman, I., Sussman, J.L., Klebe, G., Heine, A. Old acquantance rediscovered, use of xenon/protein complexes as a generic tool for SAD phasing of in house data. *To be Publ.*

70. English, a C., Done, S. H., Caves, L. S., Groom, C. R. & Hubbard, R. E. Locating interaction sites on proteins: the crystal structure of thermolysin soaked in 2% to 100% isopropanol. *Proteins* **37,** 628–40 (1999).

71. Nowak, E., Panjikar, S., Tucker, P. A. Structure of glucose isomerase derivatized with Kr. *To be Publ.*

72. Polentarutti, M., Glazer, R. & Djinović Carugo, K. A helium-purged beam path to improve soft and softer X-ray data quality. *J. Appl. Crystallogr.* **37,** 319–324 (2004).

73. Mueller-Dieckmann, C. *et al.* On the routine use of soft X-rays in macromolecular crystallography. Part II. Data-collection wavelength and scaling models. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 28–38 (2003).

74. Russo Krauss, I., Sica, F., Mattia, C. A. & Merlino, A. Increasing the X-ray diffraction power of protein crystals by dehydration: the case of bovine serum albumin and a survey of literature data. *Int. J. Mol. Sci.* **13,** 3782–800 (2012).

75. Sanchez-Weatherby, J. *et al.* Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallogr. D. Biol. Crystallogr.* **65,** 1237–46 (2009).

*Reference management handled by Microsoft Word plugin Mendeley

8.0 Appendices

Appendix A: AutoProcess results from earlier HEWL

Date Collected	#	Energy (keV)	Time (min)	Pressure (psi)	Processing Score	Mosaicity	Space Group	Cell Dimensions a	b	c	Multiplicity	Resolution (low) (Å)	Anomalous Cutoff (Å)
Jul 28 2011		7	10	400	0.74	0.09	P41212	78.9	78.9	36.8	7	2.03 (9.08)	2.74
Sep 7 2011		7	15	280	0.69	0.19	P41212	78.8	78.8	36.9	4.3	2.03 (9.08)	2.34
Jul 28 2011		7	10	200	0.74	0.18	P41212	78.9	78.9	36.8	4.2	2.03 (9.08)	2.43
Sep 7 2011		7	15	200	0.59	0.16	P41212	78.8	78.8	37	4.5	2.03 (9.06)	3.2
Sep 7 2011		7	15	160	0.64	0.14	P41212	79	79	37	4.3	2.03 (9.08)	3.21
Sep 7 2011		7	15	100	0.72	0.1	P41212	78.8	78.8	36.9	4.3	2.03 (9.08)	3.03

Date Collected	#	Energy (keV)	Time (min)	Pressure (psi)	in highest resolution anomalous shell I/σ(I)	Rmrgd	Completeness (%)	Ano Corr	Anom Signal Max	Anom Signal Total	Ano Corr Total	Notes
Jul 28 2011		7	10	400	49.18	1.8	100	65	3.082	1.649	60	
Sep 7 2011		7	15	280	35.97	2.5	100	50	2.375	1.54	56	
Jul 28 2011		7	10	200	49.62	1.8	100	59	3.13	1.749	60	
Sep 7 2011		7	15	200	52.09	1.5	100	43	2.454	1.293	38	
Sep 7 2011		7	15	160	35.83	2.1	100	49	1.935	1.244	44	
Sep 7 2011		7	15	100	41.33	1.9	100	45	2.111	1.352	49	

Appendix A: AutoProcess results for HEWL

Date Collected	#	Energy (keV)	Time (min)	Pressure (psi)	Processing Score	Mosaicity	Space Group	Cell Dimensions a	b	c	Multiplicity	Resolution (low) (Å)	Anomalous Cutoff (Å)
Nov 29 2011	1	7	Control	Control	0.59	0.25		78.1	78.1	37	5.3		
Nov 29 2011	2	7	Control	Control	0.65	0.27		78.3	78.3	37	5.2		
Nov 30 2011	3	7	Control	Control	0.58	0.4		78.2	78.2	37	6.3		
Nov 30 2011	4	7	Control	Control	0.69	0.17		78.6	78.6	36.9	7.8		
Mar14 2012	1	7	60	500	0.76	0.11		79	79	36.9	9.3	2.13 (9.04)	2.19
Mar14 2012	2	7	60	500	0.74	0.14		78.8	78.8	36.9	9.3	2.13 (9.03)	2.33
Mar14 2012	3	7	60	500	0.67	0.14		78.9	78.9	36.8	10.1	2.26 (9.03)	2.33
Mar14 2012	4	7	60	500	0.66	0.14		78.8	78.8	36.8	9.6	2.19 (9.03)	2.26
	1	7	5	350	No Crystal								
Nov 30 2011	2	7	5	350	0.64	0.23		78.8	78.8	37	8	2.14 (9.08)	2.34
Nov 30 2011	3	7	5	350	0.58	0.41		78.8	78.8	37	7.6	2.2 (9.08)	3.43
Nov 30 2011	4	7	5	350	0.54	0.51		78.9	78.9	37	7.6	2.35 (9.08)	
Dec14 2011	1	7	5	300	0.7	0.16		78.9	78.9	36.9	9.2	2.2 (9.08)	2.34
Dec14 2011	2	7	5	300	0.67	0.13	P41212	78.8	78.8	36.9	10	2.27 (9.08)	2.34
Dec14 2011	3	7	5	300	0.58	0.21		78.8	78.8	36.9	6.9	2.34 (9.08)	2.34
Dec14 2011	4	7	5	300	0.61	0.2	P41212	79.1	79.1	36.9	10.1	2.03 (9.08)	2.03
Dec14 2011	1	7	5	250	0.69	0.38		79	79	37	9	2.08 (9.08)	2.43
Dec14 2011	2	7	5	250	0.67	0.23		78.8	78.8	37	9.1	2.14 (9.06)	2.51
Dec14 2011	3	7	5	250	0.59	0.31		78.9	78.9	37	9.5	2.2 (9.08)	2.52
Dec14 2011	4	7	5	250	0.62	0.26		78.8	78.8	37	9.5	2.2 (9.08)	2.34
Jan19 2012	1	7	5	200	0.66	0.44		79	79	36.9	7.7	2.2 (9.06)	2.26
Jan19 2012	2	7	5	200	0.54	0.64		78.9	78.9	36.9	7.9	2.27 (9.08)	2.34
Jan19 2012	3	7	5	200	0.62	0.13		78.7	78.7	36.8	10.2	2.27 (9.08)	2.34
	4	7	5	200	No Crystal								
Jan27 2012	1	7	5	150	0.55	0.62		77.4	77.4	37.1	8.3	2.13 (9.30)	4.65
Jan27 2012	2	7	5	150	0.43	1.04		78.8	78.8	36.9	7.4	N/A	
Jan27 2012	3	7	5	150	0.55	0.33		78.8	78.8	36.9	9.9	2.42 (9.05)	2.42
	4	7	5	150	TWINNED								
Feb2 2012	1	7	5	100	0.71	0.1		78.9	78.9	36.9	9.4		
Feb2 2012	2	7	5	100	0.57	0.12		79.2	79.2	37	10.8		
Feb14 2012	3	7	5	100	0.58	0.13		78.9	78.9	36.9	10.9		
Feb14 2012	4	7	5	100	0.62	0.17		78.9	78.9	36.9	9.8		
Feb14 2012	1	7	5	50	0.67	0.14		78.9	78.9	36.9	9.9		
Feb14 2012	2	7	5	50	0.64	0.12		78.9	78.9	36.9	10.2		
Feb14 2012	3	7	5	50									
Feb14 2012	4	7	5	50									

Appendix A: AutoProcess results for HEWL

Date	#	Energy	Time	Pressure	in highest resolution shell				Anom Signal	Anom Signal	Ano Corr	Notes
Collected		(keV)	(min)	(psi)	I/σ(I)	Rmrgd	Completeness (%)	Ano Corr	Max	Total	Total	
Nov 29 2011	1	7	Control	Control					2.5	1.354	38	
Nov 29 2011	2	7	Control	Control					2.019	1.008	21	
Nov 30 2011	3	7	Control	Control					2.085	1.182	32	
Nov 30 2011	4	7	Control	Control					3.345	1.719	57	*
Mar14 2012	1	7	60	500	43.38	2	100	69	3.66	2.268	75	only frozen 1 hours before collection
Mar14 2012	2	7	60	500	64.95	1.4	100	76	4.153	2.281	70	collected immediately after freezing
Mar14 2012	3	7	60	500	67.12	1.3	100	81	3.893	2.581	78	collected immediately after freezing
Mar14 2012	4	7	60	500	53.47	1.8	99.9	69	3.623	2.241	69	
	1	7	5	350								
Nov 30 2011	2	7	5	350	30.46	3.3	100	53	2.905	1.564	49	
Nov 30 2011	3	7	5	350	37.94	2	100	43	2.065	1.054	19	*
Nov 30 2011	4	7	5	350					2.787	1.479	45	
Dec14 2011	1	7	5	300	47	2	100	72	3.66	2.098	66	
Dec14 2011	2	7	5	300	67.88	1.4	96.8	77	4.167	2.803	82	
Dec14 2011	3	7	5	300	41.39	2.2	97.1	58	2.95	1.804	60	
Dec14 2011	4	7	5	300	20.04	7.7	47.8	23	3.271	2.165	70	Structure built first try
Dec14 2011	1	7	5	250	33.17	3.1	100	43	2.77	1.453	48	
Dec14 2011	2	7	5	250	39	2.1	100	62	3.166	1.542	47	
Dec14 2011	3	7	5	250	32.08	3.1	100	49	2.968	1.432	40	
Dec14 2011	4	7	5	250	38.65	2.4	100	48	2.782	1.516	48	
Jan19 2012	1	7	5	200	25.57	2.3	100	61	3.799	1.467	71	
Jan19 2012	2	7	5	200	24.88	2.2	99.7	62	3.594	1.231	66	
Jan19 2012	3	7	5	200	42.43	2.5	99.9	60	3.843	1.975	59	
	4	7	5	200								
Jan27 2012	1	7	5	150	45.61	1.8	99.8	55	2.159	0.841	3	
Jan27 2012	2	7	5	150					1.244	0.791	3	
Jan27 2012	3	7	5	150	39.92	2.5	96.8	54	2.906	1.543	50	
	4	7	5	150								
Feb2 2012	1	7	5	100					3.667	1.768	52	
Feb2 2012	2	7	5	100					2.582	1.133	25	
Feb14 2012	3	7	5	100					2.036	1.504	53	
Feb14 2012	4	7	5	100					3.79	2.113	67	
Feb14 2012	1	7	5	50					3.469	1.679	54	
Feb14 2012	2	7	5	50					3.605	1.831	59	
Feb14 2012	3	7	5	50								
Feb14 2012	4	7	5	50								

Appendix A: Autoprocess results for other proteins

	Date	Puck #	#	Energy	Time	Pressure	Processing	Mosaicity	Space	Cell Dimensions			Multiplicity	Resolution (low)	Anomalous
	Collected			(keV)	(min)	(psi)	Score		Group	a	b	c		(Å)	Cutoff (Å)
GI	May 15 2012		1	7	5	500	0.6	0.21	I222	92.7	98.1	102	5	2.27 (9.08)	3.03
THERM	Jul5 2012			7	5	300	0.68	0.15	P6122	92.9	92.9	129.7	13.2	2.08 (9.08)	2.27
THERM	Oct5 2011			6.5	15	160	0.63	0.14	P622	93	93	129.3	6.4	2.44 (9.76)	2.94
THAU	May15 2012			7	5	500	0.61	0.09	P41212	57.8	57.8	150.2	10.6	2.27 (9.09)	2.74
THAU	Jul5 2012			7	5	500	0.63	0.09	P41212	57.9	57.9	150.2	11.6	2.27 (9.08)	3.03
THAU	Jul5 2012			7	5	300	0.72	0.09	P41212	57.8	57.8	150.2	10.3	2.2 (9.08)	2.27
THAU	Sep21 2011			7	15	280	0.41	0.59	P4212	58	58	150.4	3.9	2.34 (9.08)	6.42
THAU	Sep21 2011			7	15	200	0.59	0.16	P41212	58	58	150.5	4.7	2.34 (9.08)	4.06
THAU	Sep28 2011		1	6.5	15	160	0.58	0.3	P41212	58	58	150.5	3.5	2.3 (9.76)	3.98
THAU	Sep28 2011		1	6	15	160	0.59	0.3	P41212	58	58	150.5	3.5	2.49 (10.57)	5.28
THAU	Sep28 2011		1	6.5	15	100	0.69	0.14	P41212	57.9	57.9	150	3.2	2.18 (9.76)	4.37
THAU	Sep28 2011		1	6	15	100	0.7	0.11	P41212	58	58	150.4	3.2	2.37 (10.6)	3.35

Appendix A: Autoprocess results for other proteins														
	Date	Puck #	#	Energy	Time	Pressure	in highest resolution shell				Anom Signal	Anom Signal	Ano Corr	Notes
	Collected			(keV)	(min)	(psi)	I/σ(I)	Rmrgd	Completeness (%)	Ano Corr	Max	Total	Total	
GI	May 15 2012		1	7	5	500	28.99	3.8	100	42	2.087	1.133	29	
THERM	Jul5 2012			7	5	300	24.49	10.8	100	52	3.939	2.153	66	
THERM	Oct5 2011			6.5	15	160	22.46	5.3	99.5	48	2.616	1.338	39	
THAU	May15 2012			7	5	500	46.81	2.4	98.3	41	3.061	1.334	34	
THAU	Jul5 2012			7	5	500	49.2	4.7	98.6	43	2.643	1.235	36	
THAU	Jul5 2012			7	5	300	58.48	3.4	100	55	3.293	1.688	56	
THAU	Sep21 2011			7	15	280	22.27	5	99.4	45	1.239	0.82	3	
THAU	Sep21 2011			7	15	200	58.53	1.3	99.9	48	2.113	1.058	20	
THAU	Sep28 2011		1	6.5	15	160	49.72	2.1	99.3	28	2.161	1.002	12	
THAU	Sep28 2011		1	6	15	160	40.92	2.5	99.4	47	2.064	1.043	18	
THAU	Sep28 2011		1	6.5	15	100	54.65	1.5	100	49	1.88	1.089	22	
THAU	Sep28 2011		1	6	15	100	40.03	2.7	100	25	2.007	1.146	27	

Appendix B: HySS & AutoSol results for earlier lysozyme

					Hybrid Substructure Search (HYSS)					AutoSol											
					Input			Output		Input		Output									
Date	#	Energy	Time	Pressure								# Sites					Space	Xenon Sites (fractional)			
Collected		(keV)	(min)	(psi)	# Sites	High Resolution	Low Resolution	# Sites	CC	# Sites	High Resolution Limit	Refined	Rwork	Rfree	Model CC	FOM	Group	x	y	z	occ (>0.10)
Jul 28 2011		7	10	400	8	2.74	9.08	8	0.23	8	2.74	8	0.5648	0.5812	0.25	0.26	P43212	0.439	0.297	0.058	0.14
Sep 7 2011		7	15	280	16	2.34	9.08	11	0.32	11	2.34	19	0.5557	0.5952	0.29	0.418	P43212	0.547	0.283	0.046	0.13
																		0.309	0.309	0	0.11
Sep 7 2011 Same crystal		7	15	280				11	0.32	11	2.34	19	0.3737	0.4326	0.69	0.418	P43212	Same as above			
Same as above using thorough AutoSol																					
Sep 7 2011		7	15	200	19	3.2	9.08	19	0.41	19	3.2	18	0.5436	0.595	0.32	0.346	P43212	0.102	0.866	0.098	0.11
																		0.802	0.32	0.006	0.1
Jul 28 2011		7	10	200	15	2.43	9.08	15	0.26	15	2.43	15	0.5553	0.5631	0.17	0.284	P41212	-	-	-	-
Sep 7 2011		7	15	160	9	3.21	9.08	9	0.31	9	3.21	9	0.4924	0.5812	0.36	0.293	P43212	0.796	0.143	0.066	0.16
																		0.043	0.599	0.004	0.12
																		0.936	0.359	0.088	0.1
																		0.448	0.694	0.076	0.11
																		0.479	0.588	0.026	0.1
Sep 7 2011		7	15	100	19	3.03	9.08	19	0.39	19	3.03	19	0.4865	0.5987	0.39	0.342	P41212	-0.839	-0.137	-0.082	0.11
																		-0.088	-0.5	-0.099	0.1

Appendix B: HySS results for lysozyme

					Hybrid Substructure Search (HYSS)				
					Input			Output	
Date Collected	#	Energy (keV)	Time (min)	Pressure (psi)	# Sites	High Resolution	Low Resolution	# Sites	CC
Mar14 2012	1	7	60	500	13	2.19	9.04	13	0.43
Mar14 2012	2	7	60	500	15	2.33	9.03	13	0.39
Mar14 2012	3	7	60	500	20	2.33	9.03	14	0.42
Mar14 2012	4	7	60	500	20	2.26	9.03	17	0.43
	1	7	5	350					
Nov 30 2011	2	7	5	350	11	2.34	9.08	11	0.37
Nov 30 2011	3	7	5	350	14	3.43	9.08	14	0.41
Nov 30 2011	4	7	5	350	24	2.35	9.08	16	0.36
Dec14 2011	1	7	5	300	25	2.35	9.08	17	0.41
Dec14 2011	2	7	5	300	23	2.34	9.08	18	0.43
Dec14 2011	3	7	5	300	20	2.34	9.08	14	0.38
Dec14 2011	4	7	5	300					
Dec14 2011	1	7	5	250	20	2.43	9.08	16	0.37
Dec14 2011	2	7	5	250	19	2.51	9.06	13	0.41
Dec14 2011	3	7	5	250	15	2.52	9.08	9	0.33
Dec14 2011	4	7	5	250	17	2.43	9.08	13	0.42
Jan19 2012	1	7	5	200	13	2.34	9.08	13	0.24
Jan19 2012	2	7	5	200	17	2.34	9.08	15	0.41
Jan19 2012	3	7	5	200	20	2.34	9.08	14	0.38
	4	7	5	200					

					AutoSol												
					Input		Output										
Date	#	Energy	Time	Pressure			# Sites					Space	Xenon Sites (fractional)				Notes
Collected		(keV)	(min)	(psi)	# Sites	High Resolution Limit	Refined	Rwork	Rfree	Model CC	FOM	Group	x	y	z	occ (>0.10)	
Mar14 2012	1	7	60	500	13	2.13	20	0.4499	0.4806	0.5	0.447	P43212	0.454	0.718	0.043	0.15	
													0.691	0.691	0	0.1	
							20	0.4244	0.4535	0.63	0.447	P43212	0.454	0.718	0.042	0.14	thorough
													0.691	0.691	0	0.09	
Mar14 2012	2	7	60	500	13	2.13	19	0.4478	0.482	0.52	0.444	P43212	0.453	0.718	0.043	0.13	
													0.691	0.691	0	0.08	
Mar14 2012	3	7	60	500	14	2.26	14	0.495	0.5298	0.43	0.46	P43212	0.046	0.782	0.046	0.17	
													0.81	0.81	0	0.12	
Mar14 2012	4	7	60	500	17	2.19	21	0.449	0.5052	0.5	0.457	P43212	0.453	0.718	0.044	0.16	
													0.691	0.691	0	0.13	
	1	7	5	350													
Nov 30 2011	2	7	5	350	11	2.14	21	0.4441	0.4734	0.5	0.42	P43212	0.454	0.717	0.046	0.18	
													0.69	0.69	0	0.11	
Nov 30 2011	3	7	5	350	14	2.2	14	0.5592	0.5715	0.11	0.237	P41212	-	-	-	-	
Nov 30 2011	4	7	5	350	16	2.35	21	0.5413	0.5604	0.38	0.414	P43212	0.454	0.717	0.046	0.18	
													0.691	0.691	0	0.12	
Dec14 2011	1	7	5	300	17	2.2	22	0.4497	0.4787	0.5	0.452	P43212	0.455	0.717	0.047	0.21	
													0.691	0.691	0	0.12	
Dec14 2011	2	7	5	300	18	2.27	22	0.4735	0.516	0.5	0.468	P43212	0.046	0.783	0.046	0.19	
													0.809	0.809	0	0.13	
Dec14 2011	3	7	5	300	14	2.34	22	0.4921	0.5426	0.45	0.448	P43212	0.046	0.783	0.046	0.17	
													0.809	0.809	0	0.11	
Dec14 2011	4	7	5	300	2	2.03	22	0.28	0.3	0.834	0.44	P43212	0.456	0.717	0.049	0.18	Thorough build, special
													0.693	0.693	0	0.14	
													0.105	0.401	0.454	0.08	All other sites <0.7
Dec14 2011	1	7	5	250	16	2.08	21	0.4725	0.5304	0.47	0.387	P43212	0.046	0.782	0.043	0.16	
													0.809	0.809	0	0.13	
Dec14 2011	2	7	5	250	13	2.14	22	0.4471	0.4516	0.49	0.419	P43212	0.546	0.283	0.045	0.16	
													0.309	0.309	0	0.13	
Dec14 2011	3	7	5	250	9	2.2	17	0.5326	0.5661	0.36	0.385	P43212	0.045	0.782	0.047	0.18	
													0.809	0.809	0	0.13	
Dec14 2011	4	7	5	250	13	2.2	20	0.4681	0.49	0.47	0.414	P43212	0.453	0.717	0.045	0.13	
													0.691	0.691	0	0.09	
Jan19 2012	1	7	5	200	13	2.27	19	0.5209	0.547	0.37	0.42	P43212	0.283	0.544	0.047	0.23	
													0.308	0.308	0	0.11	
Jan19 2012	2	7	5	200	15	2.27	14	0.5776	0.5937	0.26	0.236	P43212	-	-	-	-	
Jan19 2012	3	7	5	200	14	2.27	20	0.4964	0.5325	0.43	0.416	P43212	0.454	0.717	0.046	0.14	
													0.69	0.69	0	0.11	
	4	7	5	200													

Appendix B: Hyss results for lysozyme

					Hybrid Substructure Search (HYSS)				
					Input			Output	
Date Collected	#	Energy (keV)	Time (min)	Pressure (psi)	# Sites	High Resolution	Low Resolution	# Sites	CC
Jan27 2012	1	7	5	150	19	4.65	9.3	19	0.67
Jan27 2012	2	7	5	150	N/A				
Jan27 2012	3	7	5	150	7	2.42	9.05	7	0.18
	4	7	5	150	N/A				
Feb2 2012	1	7	5	100					
Feb2 2012	2	7	5	100					
Feb14 2012	3	7	5	100					
Feb14 2012	4	7	5	100					
Feb14 2012	1	7	5	50					
Feb14 2012	2	7	5	50					
Feb14 2012	3	7	5	50					
Feb14 2012	4	7	5	50					
Nov 29 2011	1	7	Control	Control					
Nov 29 2011	2	7	Control	Control					
Nov 30 2011	3	7	Control	Control					
Nov 30 2011	4	7	Control	Control					

Appendix B: AutoSol results for lysozyme

					AutoSol												
					Input		Output										
Date Collected	#	Energy (keV)	Time (min)	Pressure (psi)	# Sites	High Resolution Limit	# Sites Refined	Rwork	Rfree	Model CC	FOM	Space Group	Xenon Sites (fractional) x	y	z	occ (>0.10)	Notes
Jan27 2012	1	7	5	150	19	2.13	17	0.5859	0.5979	0.08	0.216	P41212	-	-	-	-	Fast build
								0.5745	0.557	0.22	0.216	P41212	-	-	-	-	Thorough build
Jan27 2012	2	7	5	150													
Jan27 2012	3	7	5	150	7	2.42	7	0.6184	0.6208	0.24	0.209	P4212	-	-	-	-	
	4	7	5	150													
Feb2 2012	1	7	5	100													
Feb2 2012	2	7	5	100													
Feb14 2012	3	7	5	100													
Feb14 2012	4	7	5	100													
Feb14 2012	1	7	5	50													
Feb14 2012	2	7	5	50													
Feb14 2012	3	7	5	50													
Feb14 2012	4	7	5	50													
Nov 29 2011	1	7	Control	Control													
Nov 29 2011	2	7	Control	Control													
Nov 30 2011	3	7	Control	Control													
Nov 30 2011	4	7	Control	Control													

Appendix B: HySS & Autosol results for other proteins

							Hybrid Substructure Search (HYSS)					AutoSol												
							Input			Output		Input		Output										
	Date	Puck #	#	Energy	Time	Pressure								# Sites					Space	Xenon Sites (fractional)				
	Collected			(keV)	(min)	(psi)	# Sites	High Resolution	Low Resolution	# Sites	CC	# Sites	High Resolution Limit	Refined	Rwork	Rfree	Model CC	FOM	Group	x	y	z	occ (>0.10)	
GI	May 15 2012		1	7	5	500	19	3.03	9.08	14	0.4	14	2.27	11	0.4035	0.4259	0.55	0.371	I222	-0.428	0.271	-0.357	0.59	Magnesium??
THERM	Jul5 2012			7	5	300	9	2.27	9.08	9	0.55	9	2.08	8	0.3999	0.443	0.58	0.475	P6122	-0.042	-0.347	-0.625	0.8	
THERM	Oct 5 2011			6.5	15	160	20	2.94	9.76	17	0.26	17	2.44	14	0.5416	0.5342	0.28	0.424	P622	0	0	0	2.36	
THAU	May15 2012			7	5	500	19	2.74	9.09	9	0.41	9	2.27	17	0.433	0.4606	0.56	0.337	P41212	-0.07	-0.219	-0.052	0.1	
THAU	Jul5 2012			7	5	500	14	3.03	9.08	9	0.32	9	2.27	18	0.4372	0.4606	0.51	0.338	P41212	-0.07	-0.22	-0.051	0.1	
THAU	Jul5 2012			7	5	300	13	2.27	9.08	10	0.27	10	2.2	21	0.4183	0.4516	0.55	0.429	P41212	-0.076	-0.219	-0.057	0.13	
THAU	Sep21 2011			7	15	280	13	6.42	9.08	9	0.59	9	2.34	7	-	-	-	0.158	P4212	-	-	-	-	
THAU	Sep21 2011			7	15	200	9	4.06	9.08	9	0.33	9	2.34	8	0.5738	0.5623	0.19	0.172	P43212	-	-	-	-	
THAU	Sep28 2011		1	6.5	15	160	19	3.98	9.76	19	0.44	19	2.3	17	0.5879	0.5805	0.12	0.201	P43212	-	-	-	-	
THAU	Sep28 2011		1	6	15	160	16	5.28	10.57	16	0.52	16	2.49	15	0.5659	0.5657	0.07	0.2	P41212	-	-	-	-	
THAU	Sep28 2011		1	6.5	15	100	14	4.37	9.76	14	0.42	14	2.18	14	0.5668	0.5857	0.09	0.175	P41212	-	-	-	-	
THAU	Sep28 2011		1	6	15	100	24	3.35	10.6	16	0.27	16	2.37	16	0.5637	0.5808	0.14	0.201	P43212	0.147	0.166	0.105	0.13	